2016

Effects of HIV Behavioral Interventions for Men Who Have Sex With Men - United States, 1988-2014

H. Elsa Larson
University of Rhode Island, helarson@my.uri.edu

EFFECTS OF HIV BEHAVIORAL INTERVENTIONS FOR MEN WHO HAVE SEX WITH MEN – UNITED STATES, 1988-2014

BY

H. ELSA LARSON

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2016

DOCTOR OF PHILOSOPHY DISSERTATION

OF

HEIDI ELSA LARSON

APPROVED:

Dissertation Committee:

Major Professor: Joseph S. Rossi

Colleen A. Redding

Aisling R. Caffrey

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2016

ABSTRACT

Men who have sex with men (MSM) in the United States continue to experience disproportionate HIV disease burden. Historically, most HIV prevention research has focused on reducing sexual risk behaviors through individual-level behavior change interventions. To date, behavioral interventions have not reduced HIV incidence among MSM and combination prevention approaches that package behavioral interventions with additive biomedical or structural interventions are now recommended. The last meta-analysis of HIV behavioral interventions for MSM was conducted in 2008. Since then sixteen new rigorous trials have been identified. New evidence and new recommendations justify an updated meta-analysis to identify the most promising and relevant features of behavioral interventions to be used in new combination approaches. This study aimed to calculate an updated effect size for MSM-specific HIV behavioral interventions, identify moderators of effect size, examine cumulative effect sizes over time, and describe trials that addressed more than one behavioral outcome ("integrated interventions"). Systematic review and meta-analysis evaluated effects of 34 randomized controlled trials for 17,872 US MSM conducted between 1989 and 2014. Behavioral interventions reduced the odds of sexual risk behavior by 14 percent (*OR*=.859, 95% CI [0.790, 0.933], $p<.001$). Findings suggest behavioral interventions are still somewhat effective to reduce sexual risk behavior, but the effect size was smaller than effect sizes observed in earlier meta-analyses. Cumulative meta-analysis further demonstrated that intervention effects gradually declined over time. From 1991 to 2014, the magnitude of the effect size decreased by 19.5 percent (*OR*=.719-.859). Reasons for effect size decline are not

clear, but HIV prevention fatigue, inclusion criteria that focus on very high-risk MSM, choice of comparison condition, and underpowered primary trials likely contribute to effect size shrinkage. Statistical homogeneity restricted this study's objective to reliably detect moderators of effect [$Q(33)=39.35$, $p=.207$; $I^2=16.14$]. All moderators hypothesized *a priori* were not significant. Post-hoc moderator analyses found intervention effects to be moderated by age ($p<.001$), peer delivery ($p=.002$), community-level interventions ($p=.032$), HIV status ($p=.019$), education ($p=.023$), evidence-level ($p=.076$), retention ($p=.09$), and MSM subgroup ($p=.09$). Nine trials were identified that addressed at least one additional problem behavior other sexual risk behavior; six trials addressed substance use and three trials addressed HIV testing. In conclusion, this study provides new evidence that behavioral interventions have become less effective over time. Development of new combination prevention packages presents an opportune time to improve and update behavioral interventions. HIV prevention research would benefit from frequent research synthesis to monitor effect sizes, identify the most effective intervention components, retire outdated intervention components, and identify gaps in current research. To our knowledge, this is the first meta-analysis of HIV behavioral interventions to demonstrate effect size shrinkage for HIV behavioral interventions using cumulative meta-analysis. Routine cumulative meta-analysis should be included in research synthesis protocols to examine and explain effect size shift as new evidence is accumulated.

ACKNOWLEDGEMENTS

I acknowledge the entire dissertation committee for their time, critical feedback, and motivational support. I thank Dr. Susan Rossi for her willingness to serve as my committee chair, and for her overall kindness and generous spirit. I thank Dr. Andrea Paiva for her ongoing participation in my committees, and for sharing her keen analytical skills. Dr. Paiva was a member of my pharmacoepidemiology masters committee where she provided substantive and methodological expertise about examining predictors of non-vaccination for HPV among US women aged 18-26. I was also fortunate to have Dr. Aisling Caffrey as part of that committee, and I am grateful to have an epidemiologist's perspective on my dissertation committee. I thank her and Dr. Cynthia Willey for providing excellent training in the principles of epidemiology through URI's Pharmacoepidemiology program. I thank Dr. Colleen Redding for her continued support to my committees, and for sharing her vast expertise in HIV prevention and rigorous health behavior intervention research. I additionally thank her for her longstanding commitment to advancing theoretical HIV prevention research, and for her many original contributions to the HIV prevention evidence base. Finally, I thank my major professor, Dr. Joseph Rossi. I thank Dr. Rossi for providing me with seven years of patient mentoring, for sharing his broad statistical, measurement, and meta-analytic wizardry, as well as his pioneering work in power analysis. I also thank him for introducing me to the magic of Turley zinfandels and monkfish liver (ankimo) – and all those wonderful sushi dinners.

TABLE OF CONTENTS

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

FIGURE **PAGE**

CHAPTER 1: INTRODUCTION

Statement of the Problem

The first United States (US) cases of human immunodeficiency virus (HIV) were reported in 1981. Within a few years, about 50 percent of gay men living in San Francisco were infected with HIV (Wohlfeiler & Ellen, 2007). There were no HIV medications, and no government-funded HIV prevention programs. Communities mobilized through grassroots organizing to distribute information and educational materials to gay men. Pamphlets, brochures, and media explained that HIV was a preventable, sexually transmitted infection. Social norms changed in the gay community and men modified their sexual practices to avoid HIV infection. Widespread reductions in sexual risk behavior curbed HIV incidence and rates plummeted (Wohlfeiler & Ellen, 2007).

The majority of HIV prevention research since has focused on reducing sexual risk behavior to replicate the community-level behavior change seen in the 1980s. However, behavior change associated with the early HIV crisis slowly eroded over time, and it is unrealistic to expect similar effects in today's mature epidemic (Wohlfeiler & Ellen, 2007). While HIV behavioral interventions have demonstrated efficacy in reducing sexual risk behavior in small groups of gay, bisexual, and other men who have sex with men (referred to as men who have sex with men [MSM] in CDC surveillance systems), effects of interventions are limited by reach and dose and are not sustained over time. To date, no funded HIV behavioral intervention has reduced HIV incidence among US MSM. Behavioral interventions, while still

necessary, are no longer sufficient for MSM (Coates, 2013). As the HIV epidemic among US MSM becomes more complex and challenging to control, additional strategies that can enhance individual-level behavior change interventions are needed.

Combination HIV prevention packages that benefit from combining partially-effective behavioral, biomedical, and structural interventions are now recommended. Combination approaches aim to identify the most effective components of each intervention and rationally combine them. Scale up of combination approaches presents an opportune time to review the state of the science of behavioral interventions and identify the most promising and relevant intervention features. The last meta-analysis of HIV behavioral interventions for MSM was published in 2008 (Johnson et al., 2008). Since 2008, sixteen rigorous intervention trials have been published. An updated systematic review was published in 2013 (Higa et al., 2013), but did not include a meta-analysis. An updated systematic review with meta-analysis to calculate an updated effect size and identify factors associated with effectiveness is needed to inform the development of new behavioral interventions.

Epidemiology of HIV among US MSM

Prevention and control of HIV continues to be a major public health challenge in the United States (Frieden, Foti, and Mermin, 2015). The Centers for Disease Control and Prevention (CDC) estimates 1.1 million people are infected with HIV (CDC, 2013a). About 1 in 6 of people infected with HIV do not know they are infected, and are at higher risk of transmitting their infection to others (CDC, 2013a; Marks, Crepaz, & Janssen, 2006). HIV incidence has been stable at about 50,000 new

HIV infections each year (CDC, 2012). However, US MSM continue to experience disproportionate HIV disease burden compared to other risk groups.

In 2010, 63 percent all persons diagnosed with HIV were MSM (CDC, 2012). MSM make up about 4 percent of males in the US, but accounted for 78 percent of all males diagnosed with HIV (CDC, 2012; CDC 2014a). MSM are 44 times more likely to be infected with HIV than heterosexual males (Purcell et al., 2012). Further, MSM is the only US risk group where new infections have increased. From 2008 to 2010, new infections among MSM significantly increased by 12 percent with the steepest increase (22 percent) among younger MSM aged 13-24 years. Young black MSM comprise more new HIV infections than any other demographic group (CDC, 2014a), highlighting a complex epidemic fueled by racial and economic disparities.

MSM of all races and ethnicities also have poorer health outcomes at each stage of HIV care (e.g. HIV testing, linkage to care, retention in care, adherence to HIV medications, and viral suppression). Poor care outcomes result in greater HIV risk because MSM may be unaware of their infection and/or have viral loads high enough for transmission to sexual partners. In 2010, 78 percent of MSM diagnosed with HIV were linked to care, 51 percent were retained in HIV care, 49 percent received antiretroviral therapy, and only 42 percent achieved viral suppression (Singh et al., 2014). In a 2011 study of MSM in 21 major US cities, 18 percent of MSM were found to be infected with HIV. One-third of HIV-positive MSM did not know they were infected (Wejnert et al., 2013).

Complex HIV Risk Environment for MSM

Since the beginning of the epidemic, HIV prevention research has been challenged to demonstrate highly effective interventions for US MSM. The social context can quickly evolve making previously effective interventions less relevant. Additionally, MSM-specific HIV epidemics differ from other groups due to multi-factorial risk environments. Biological, social, and structural factors interact to create transmission dynamics that accelerate rapid and efficient transmission of HIV through sexual networks (Beyrer et al, 2012; Johnson et al., 2010; Sullivan et al., 2012). Disproportionate HIV disease burden is driven primarily by the high per-act transmission probability of HIV infection from each encounter of unprotected anal sex (Baggeley, White, & Boiley, 2010; Beyrer, 2012; Vittinghoff et al., 1999; Patel et al., 2014). In other words, higher background prevalence of HIV increases the likelihood that sexual risk behavior will result in infection among MSM (CDC, 2014a).

Higher biological risk is further compounded by person-level factors such as having multiple sexual partners, frequency of unprotected anal sex, substance use prior to or before sex, and incident sexually transmitted infection (STI) (Koblin et al., 2006). Other factors that influence HIV transmission include social or structural determinants that impact the environment people live in. Stigma and homophobia, gaps in comprehensive health care, and poor access to condoms and/or HIV testing challenge HIV prevention norms and create additional stressors for vulnerable MSM (Altman et al., 2012; Kaufman, Cornish, Zimmerman, & Johnson, 2014; Mayer et al., 2012). Finally, community-level factors such as higher community viral load due to undiagnosed HIV infection or poor adherence to HIV medications can vary by sexual

network (e.g. male sex workers, homeless MSM, substance-using MSM), and make some communities more vulnerable to HIV exposure and infection (Das et al., 2010; Mayer et al., 2014).

Combination Prevention Interventions

HIV risk and transmission dynamics among MSM may be too complex to be effectively addressed by behavioral interventions alone. MSM who intentionally engage in high-risk sexual practice (despite knowledge of the risks) may require additional strategies including, but not limited to, individual behavior change. While newer biomedical interventions are likely to be more effective than behavioral interventions, they also are not sufficient to prevent HIV if implemented without behavioral interventions (Coates, 2013). "Combination prevention interventions" are assumed to maximize prevention through the additive effects of partially-effective behavioral, biomedical, and structural interventions (Cohen et al., 2013; Dieffenbach & Fauci, 2011; Kurth, Celum, Baeten, Vermund, & Wasserheit, 2011 Sullivan et al., 2012;).

UNAIDS (2010) defines combination HIV prevention as:

> "The strategic, simultaneous use of different classes of prevention activities (biomedical, behavioral, social/structural) that operate on multiple levels (individual, relationship, community, societal), to respond to the specific needs of particular audiences and modes of HIV transmission, and to make efficient use of resources through prioritizing, partnership, and engagement of affected communities."

Combination approaches are now recommended in the National HIV/AIDS Strategy and in the President's Emergency Plan for AIDS Relief (US Department of State, 2013; The White House, 2010). To advance new combination approaches, the most effective components of currently available biomedical, structural, and behavioral interventions need to be identified. Combination prevention packages can then be tailored to disrupt the primary drivers of HIV infection in specific populations or sexual networks.

Biomedical Interventions

Biomedical interventions for MSM take advantage of scientific advances using antiretroviral therapy (ART) as prevention and hold high promise to curb HIV transmission. Biomedical interventions, if scaled up, are estimated to have greater impact on HIV incidence than behavioral interventions - especially for high-risk MSM resistant to individual behavior change. Recommended biomedical interventions for US MSM include pre- and post-exposure prophylaxis (PrEP and PEP, respectively) for uninfected MSM, initiation of ART as "treatment as prevention" (TasP) for HIV-positive MSM, as well as expedited treatment of incident STI. PrEP and TasP are the most promising biomedical strategies based on estimated efficacy, however they are relatively recent interventions and are expected to require behavioral components to optimize adherence, and decrease sexual risk (Cohen et al., 2010; Cohen, et al., 2013; Grant et al., 2010).

Structural Interventions

Structural interventions aim to influence the environmental variables associated with risk and create environments that promote HIV prevention norms

(Kaufman et al., 2014). There is less evidence for these interventions due to challenges in evaluating slow, structural-level change, yet their hypothesized contribution to health equity and disruption of key social drivers of HIV risk support their role in combination approaches (Grossman, Purcell, Rotherum-Borus, & Veniegas, 2011; Kaufman, et al., 2014; Kurth, Celum, Baeten, Vermund, & Wasserheit, 2011). Examples of structural interventions include condom distribution programs, syringe exchange programs, comprehensive health care, and policy changes to reduce institutional discrimination against lesbian, gay, bisexual, and transgendered individuals (Breyer et al., 2012; Kaufman et al., 2014).

Behavioral Interventions

This study focuses exclusively on the state of the science of behavioral interventions. Behavioral interventions have the strongest evidence base for reducing sexual risk behavior, but they have not shown evidence for reducing incidence (Sullivan et al., 2012). However, it is logical to continue to support behavioral interventions because HIV transmission among MSM is driven primarily by the *risk behavior* of unprotected anal intercourse (UAI). UAI with serodiscordant partners (known or unknown HIV-positive partners) and subsequent exposure to high HIV viral load increases risk of HIV acquisition (Coates, Richter, & Caceres, 2008; Sullivan et al., 2012). Among US MSM, encounters of UAI increased nearly 20 percent from 2005 to 2011 (Frieden, Foti, & Mermin, 2015; Paz-Bailey et al. 2013). Behavioral risk data highlight the importance of efficacious sexual risk reduction interventions that have specific prevention targets such as: 1) reduced UAI, 2) increased condom use, 3) partner selection based on HIV status, and 4) decreased viral

load for HIV-positive MSM. Substance use prior to sex, unknown HIV status, incident STI, and poor treatment adherence among HIV-positive MSM are additional potential targets for comprehensive or "integrated" behavioral interventions (Coates et al., 2008; Grossman et al., 2011; Kurth et al., 2011; Mayer et al., 2014; Sullivan, 2012).

Efficacy of Behavioral Interventions

High-quality meta-analytic reviews conducted so far have provided reliable evidence that behavioral interventions are efficacious in reducing HIV risk among MSM either by increasing condom use (range: 61% - 81% increase) or decreasing incidence of UAI (range: 23% - 43% decrease) (Herbst, et al., 2005; Johnson et al., 2002a; Johnson et al., 2005; Johnson et al., 2008; Sullivan et al, 2012). Intervention research meta-analyses were further examined by Noar (2008) in a meta-review; the weighted mean effect sizes for UAI among MSM were consistent with previous reviews (OR range: 0.65-0.78).

Reviews identified study moderators that produced more favorable effects in sexual risk reduction interventions. Design variables were shown to most influence effects. For example, Johnson (2002a, 2008) found count outcomes to be a more sensitive outcome than dichotomous outcomes because they identify smaller, and meaningful, reductions in risk behavior. Johnson (2008) also found greater effects in group-level interventions were associated with shorter intervention spans, better retention in the intervention condition, and comparison groups with little to no HIV prevention intervention. Johnson found community-level interventions that addressed personal skill development were associated with the greatest reductions of UAI, and

that effects were associated with random assignment over convenience, shorter recall periods with longer follow-up periods, higher percentage of non-gay identified MSM, and higher percentages of white MSM. Herbst et al (2007) found individual-level interventions based on multiple health behavior theories were associated with greater effects on sexual risk reduction. Interpersonal skills, awareness, risk and loss perceptions, and self-efficacy were shown to be important moderators for group-level interventions, as well as for interventions with multiple sessions and mixed components (Herbst et al., 2005, 2007; Johnson et al., 2002a, 2007). Theory-based interventions were more effective than non-theory based interventions (Herbst et al., 2005; Noar, 2008).

Behavioral interventions are currently supported by the Centers for Disease Control and Prevention, and evidence-based interventions (EBIs) are published in the Compendium of Effective Behavioral Interventions (CDC, 2014b). CDC relies on standardized evaluation of all published interventions using CDC-designed efficacy criteria (CDC, 2014c). Additionally, CDC routinely publishes research syntheses of behavioral interventions (Johnson et al., 2002; Johnson et al., 2008; Higa et al., 2013). All reviews published to date were conducted before recommendations for combination approaches. Recommendations for combination agendas are driven by increased evidence pointing to the limitations of behavioral interventions. No review has examined behavioral interventions in the context of a combination prevention agenda, and no review has examined how effect sizes may have changed over the past decade as HIV prevention research has evolved to focus more exclusively on higher-risk MSM subpopulations.

Integrated Behavioral Interventions

This study acknowledges that behavioral interventions are necessary, but not sufficient for HIV prevention. Behavioral interventions can better contribute to combination prevention agendas if their most effective components are well-understood and applied most effectively to the highest risk populations. Higher risk populations include MSM subgroups identified to have higher burden of HIV and complex risk factors (e.g. young MSM, black MSM, substance-using MSM). Further, behavioral interventions may have more impact if they strategically target additional behavioral outcomes other than sexual risk, otherwise known as syndemic risk factors. A new area of research is exploring the prevention effects of "integrated interventions" that address multiple problem behaviors simultaneously (i.e. substance use, infrequent HIV testing) (Collins, 2015; Crepaz et al., 2014). For example, a meta-analysis of integrated interventions for HIV-positive persons demonstrated positive effects for sexual risk reduction, and promising effects for medication adherence, lending strong evidence for further research into behavioral interventions that can influence more than one problem behavior at a time (Crepaz et al., 2014).

HIV-negative MSM may also benefit from this approach. The majority of evidence to date supports only single-behavior sexual risk behavior trials. Measurement of multiple outcome indicators is not new to the field, and many trials already include two to three indicators of sexual risk (i.e. UAI, condom use, number of partners). However, very few trials in the past 25 years have measured convergent behaviors. Recently, intervention research for HIV-negative MSM has tested interventions that aim to concurrently reduce sexual and drug risk behaviors for

substance users. A recent behavioral intervention using a randomized controlled trial (RCT) design for young substance-using MSM showed positive effects for substance use and sexual risk and was classified as an evidence-based intervention by the Centers for Disease Control (Parsons et al., 2014). Research compiled from other trials suggests motivational interviewing, cognitive behavioral counseling (CBC), personalized cognitive counseling (PCC), and empowerment theory influence intervention effectiveness for substance-using MSM (Kurtz, Stall, Buttram, Surratt, & Chen, 2013; Morgenstern et al., 2012; Santos et al., 2014; Velasquez et al., 2009). However, other RCTs found no effect for PCC or CBC (Mansergh et al., 2010; Schwarcz et al., 2013), and highlight the need for updated research syntheses to clarify what works best for substance-using MSM (Melendez-Torres & Bonell, 2013).

Increasing HIV testing among high-risk populations, particularly for MSM of color, is another outcome directly related to improved prevention and care outcomes. There is less meta-analytic and experimental research specific to increasing HIV testing. Most studies lack rigor and use observational designs, or have insufficient sample sizes of MSM (Johnson et al., 2002a). It is known that MSM do not test for HIV as frequently as recommended, and this observation has stimulated new and stronger research in this area (Maulsby et al., 2013; Paz-Bailey et al., 2013; Wejnert et al., 2013). HIV testing is recommended for all MSM at least once per year, but only 67 percent of US MSM reported receipt of an HIV test in the past 12 months (Paz-Bailey et al., 2013). Meta-analytic research demonstrates that knowledge of HIV status is a robust predictor of reduced risk behavior and increased disclosure to sexual partners (Marks, Crepaz, & Janssen, 2006). Individuals who test frequently are more

likely to identify acute HIV infection earlier, reduce transmission to others, and expedite linkage to care (van den Berg, Larson, Zimet & Lally, 2014). Social network-based strategies have been evaluated in primary studies among black MSM, but findings and designs have varied (Baytop et al., 2014; Ellen et al., 2013; Fuqua et al., 2011; Halkitis et al., 2011; McCree et al., 2013). Additionally, internet-based interventions show some evidence for increasing HIV testing among MSM, yet the research body is too small to be rigorously evaluated (Rhodes et al., 2011; Schnall, Travers, Rojas & Carballo-Diequez, 2014). Increasing access to convenient, free HIV testing is a current CDC priority for structural interventions and required through CDC cooperative agreements with state health departments (CDC, 2015a), but behavioral interventions will be required to increase individual motivation and receipt of HIV testing among MSM.

CHAPTER 2: STUDY JUSTIFICATION

Study Objectives

Behavioral HIV prevention interventions for MSM are efficacious in reducing sexual risk behaviors and should continue to be supported and developed. However, behavioral interventions alone are not expected to reduce HIV incidence. Research suggests that combination approaches that maximize prevention impact of all available components to prevent HIV are most promising for disrupting transmission dynamics in MSM epidemics. New behavioral interventions used in combination approaches need to be based on the most current evidence, relevant to today's mature epidemic, and effective with high-risk MSM. This study aims to identify the most promising and relevant features of behavioral interventions for MSM to inform the development of new behavioral interventions.

Objectives of this study are to:

1. Locate and describe experimental outcome studies evaluating effects of behavioral HIV interventions for US MSM.
2. Summarize effectiveness of interventions to reduce sexual risk behavior using meta-analysis.
3. Identify study moderators associated with effectiveness.
4. Display cumulative effect sizes over time.
5. Identify experimental trials that examine multiple problem behaviors in addition to sexual risk reduction such as increased HIV testing and decreased substance use.

Study Justification

HIV prevention for US MSM is a persistent public health challenge and new approaches are urgently needed. Trials testing new combination prevention approaches are currently underway (e.g. HIV Prevention Trials Network 080 and 073). To support the role of behavioral interventions in new combination agendas, updated research syntheses are needed to re-evaluate the current overall effect size, examine shifts in effect size over time, and identify other behavioral outcomes feasible for "integrated interventions."

The last research synthesis was conducted by the Prevention Research Synthesis Team at the Centers for Disease Control (CDC). Their team published the last meta-analysis of findings in 2008 (including trials up to 2007) by Johnson et al. This review included all known RCT designs testing any behavioral intervention that included some proportion of MSM participants. This review included trials inside and outside of the US and did not require trials to be specifically designed for MSM (e.g. trials designed for HIV-positive clinic populations generally have higher proportions of MSM, but MSM were not the trial focus). In 2013, the team performed a (mostly) qualitative synthesis of interventions previously evaluated by CDC to meet criteria for being specifically designed for US MSM (Higa et al., 2013). This systematic review revealed trends in HIV prevention research and helped explain reasons for failed efficacy. However, it did not quantitatively summarize findings to produce an updated effect size. While systematic reviews are essential to research synthesis, they are limited if they do not calculate updated effect sizes. Quantitative analyses, or meta-analytic results, are required to justify the rationale for ongoing behavioral research in

MSM, focus research questions based on reliable data, set benchmarks to validate future research, and examine moderators of effect (Cook et al., 1992; Marsh, Johnson, & Carey, 2001; Johnson et al., 2008).

Study Hypotheses

This study used systematic review and meta-analysis to synthesize sexual risk reduction behavioral interventions specifically designed for adult US MSM from 1988-2014. Eligible trials that include integrated interventions such as HIV testing and substance use were summarized to evaluate the prevalence of integrated interventions. Hypotheses for this study were:

1. Behavioral interventions are effective to reduce sexual risk behavior among US MSM.
2. Effects of behavioral interventions are moderated by design variables (i.e. outcome measure, comparison condition, differential retention, and intervention time span).
3. An updated meta-analysis will validate effect sizes observed in previous reviews.
4. Integrated interventions, or trials with multiple behavioral outcomes, have increased since the last review.

CHAPTER 3: Methodology

Systematic Review

A systematic review of the HIV prevention literature was conducted to locate and retrieve trials evaluating behavioral interventions for US MSM and published between 1988 and 2014. The systematic review procedure was informed by guidelines from *The Cochrane Handbook for Systematic Reviews of Interventions* (Higgins & Green, 2008) and Preferred Reporting Guidelines for Systematic Review and Meta-Analysis (Moher, Liberati, Tetzlaff, Altman & PRISMA Group, 2009).

The review was influenced by previous high-quality meta-analyses that have been used to inform evidence-based recommendations for HIV prevention (Herbst et al., 2007; Johnson et al., 2008; Higa et al 2013). The present review differs from other reviews because it excludes interventions designed exclusively for HIV-positive MSM, and it only includes trials specifically designed for MSM. Trials specific to HIV-positive MSM did not fall under this study's classification as HIV prevention to prevent HIV acquisition. Following Higa et al.'s approach, only trials designed for MSM were included to understand intervention effects and past research specific to this population. This review updates the qualitative synthesis of Higa et al. (2013) by adding randomized controlled trials from 2011-2014, and also providing a meta-analytic synthesis of trials published from 1988-2014.

Inclusion Criteria

Trials were reviewed for relevance based on four major criteria: 1) types of participants (i.e., MSM), 2) types of outcome measures (i.e., HIV risk behaviors), 3) types of interventions (i.e., behavioral interventions), and 4) types of trials (i.e., RCT

design). Trials were considered in scope if they examined any behavioral intervention aimed at reducing sexual risk behaviors for HIV transmission among US MSM. Trials were eligible for inclusion if they were: 1) behavioral interventions to prevent acquisition of HIV, 2) specifically designed for adult MSM, 3) conducted in the United States, 4) tested using a randomized controlled trial, 5) measured at least one behavioral or biological outcome variable of relevance (e.g., HIV or STI incidence, anal sex without a condom, number of sexual partners, condom use for anal sex), and 6) published in a peer-reviewed journal between 1988 and 2014.

Exclusion Criteria

Review was restricted to interventions focused on the primary prevention of HIV infection among adult MSM. This restriction assumes there may be important differences between HIV-negative MSM and HIV-positive MSM, as well as differences between adult MSM and adolescent MSM, in the underlying processes of behavior change (Herbst et al., 2007). Inclusion of such trials may have increased statistical heterogeneity or obscured effects of prevention interventions on the population of interest: HIV-negative adult MSM. Trials were excluded if they: 1) focused on HIV-positive MSM (i.e., 100 percent HIV-positive sample) and secondary prevention of HIV transmission to sexual partners, 2) included some proportion of MSM participants, but were not specifically designed for adult MSM (i.e., less than 95 percent MSM), and 3) focused on adolescent MSM (i.e., mean sample age of <18 years). Trials that did not report enough statistical information to calculate an effect size were excluded after attempts to contact study authors failed to produce necessary data.

Types of Participants

Trials were included if participants were adult men over 18 years old who reported sex with other men and lived in the United States. MSM were included regardless of race/ethnicity, sexual identity (e.g., gay, bisexual, homosexual, heterosexual, etc.), or other demographic characteristics. Trials that included some proportion of HIV-positive MSM in the sample were included except when the trial exclusively focused on HIV-positive MSM.

Types of Interventions

Trials were included if they were behavioral interventions designed to prevent the acquisition of HIV infection by changing individual sexual risk behaviors through modeling, demonstration, role-playing, risk reduction planning, group or individual counseling, or other behavioral intervention method. Individual-level, group-level, couples-level, and community-level interventions were included. Behavioral interventions were defined to be different than provision-of-information-only interventions that aim to change knowledge, attitudes, or norms only (e.g., increasing HIV knowledge), and different than environmental or structural interventions that aim to change the physical or social environment to promote health and prevent disease (e.g., mass media campaigns, policy change) (The Community Guide, 2008). Trials that focused only on psychological moderators of risk behavior, cognitive outcomes, or affective outcomes (e.g., distress, depression) were considered out of scope.

Types of Outcome Measures

Review was restricted to trials that measured intervention effects on behaviors known to influence risk of HIV acquisition among MSM (e.g., anal sex without a

condom, number of male sexual partners, frequency of condom use for anal sex), or trials with biological outcomes such as HIV or STI incidence. Operational definitions of primary outcome variables are not standardized among HIV prevention trials (Johnson et al., 2002b). The most frequently reported measures used for sexual risk behavior in most at-risk populations are frequency of unprotected sex, condom use, and number of sex partners (Johnson et al., 2002b). Unprotected sex is often reported using dichotomous (i.e., proportion of participants reporting any unprotected sex within the recall period) and count measures (i.e., number of partners, or number of episodes for unprotected sex within the recall period). Trials were eligible for review if they reported at least one measure of sexual risk behavior that was specific to male-to-male sexual contact.

Types of Trials

Trials were eligible if they included a relevant outcome variable and methodological rigor as demonstrated by study design. Only trials that used a randomized controlled trial (RCT) with independent comparison group were included for review. No exclusions were made by type of comparison group. All other study designs, including rigorous quasi-experimental, were excluded. This restriction was intended to ensure that the evidence under review had a high level of methodological rigor (Guyatt et al., 1995; Stephenson & Irmie, 1998). RCT designs are experimental studies where subjects are randomly allocated to intervention (i.e., treatment) or comparison (i.e., control) group, and then followed under controlled conditions (CEBR, 2015). When implemented with high quality, RCTs are better able than other designs to attribute observed effects to the intervention condition while also

minimizing potential sources of bias (Higa et al., 2013; CEBR, 2015). Most interventions classified as Evidence-Based Interventions (EBI) by CDC are tested using RCT design to ensure internal validity before disseminating evidence-based interventions into real-world settings.

Search Strategy

The peer-reviewed published HIV prevention literature was systematically searched from April 2015 to June 2015 to locate trials that met eligibility criteria. Five electronic databases were searched from January 1988 to April 2015: 1) PubMed, 2) EMBASE, 3) CENTRAL, 4) PsycInfo, and 5) CINAHL. The year of 1988 was chosen as the start date to capture the earliest known intervention trials and is consistent with previous reviews (Johnson et al., 2008). Search sensitivity was prioritized over precision; no restrictions by country, geography, outcomes, or language were applied. After electronic searches, hand searches of five key journals (*AIDS & Behavior, American Journal of Public Health, AIDS Education & Prevention, Journal of the Association of Nurses in AIDS Care*, and *AIDS Care*) were conducted to locate articles published between January 2014 to December 2014 that may have been missed by electronic searches due to indexing lags. References from prior systematic reviews and meta-analyses were reviewed until no new references were identified.

Keywords, search strings, and search strategies varied by electronic database. Three search strings were developed to systematically search the literature: 1) MSM string (the population filter), 2) HIV/AIDS prevention string (the disease filter), and 3) RCT string (design filter). The population filter and disease filter were specified using

controlled vocabulary (e.g., MeSH terms for PubMed or EMTREE terms for EMBASE) to retrieve articles that may be indexed by different words. Controlled vocabulary terms were identified through a comprehensive approach. Search strategies from previous reviews and consultations with experts trained in Cochrane Systematic Reviews helped to inform the first draft of each string. Strings were then refined by identifying trials known to fit eligibility criteria and looking up their index terms, common text words, and subject words. Other terms were identified using search tools in each database (e.g., MeSH database in PubMed) to customize the search to each database. Once key terms were identified, terms were exploded to generate more specific terms or synonyms to include in each search string. Boolean operators joined together controlled vocabulary terms, free-text terms, and identified synonyms for the final string. The published "Cochrane Highly Sensitive Search Strategy for Identifying Randomized Trials in Medline" (Higgins & Green, 2008) was used in its exact form to filter trials by RCT design in PubMed. This string was adapted for use in other databases. A detailed Search Strategy document (Appendix A) details the search strategy, final strings, date and time of search, results, and final number of trials located for each database searched. Citations and abstracts for located trials were downloaded from each database and exported into an EndNote X7 file for reference management. Duplicates were deleted. Titles and abstracts were scanned by the primary author to validate they met inclusion criteria. Trials that did not meet inclusion criteria were discarded into an Excluded folder. Full manuscripts that met inclusion criteria were downloaded for data extraction.

Data Extraction

A study protocol was developed to guide data extraction (Appendix B) and was informed by The Community Guide's Data Abstraction Form (2008). Data extraction forms collected a wide range of information on study descriptors (e.g., participant characteristics, study characteristics, design characteristics), study results (e.g., primary and secondary outcomes, significance, study retention), statistical information (e.g., descriptive data to calculate effect sizes), and methodological quality (e.g., CDC's Prevention Research Synthesis criteria). Two independent coders extracted trial data using standardized forms; one coder completed 50 percent of records. Disagreements were resolved by discussion until consensus was reached; a methodological expert was consulted as needed.

Data to Test Study Hypotheses

Moderating variables specified *a priori* were extracted to test study hypotheses. Based on Johnson et al. (2008), four variables were selected and coded for moderator analyses: 1) effect measure type, 2) intervention time span, 3) differential retention by group, and 4) control condition. A second hypothesis specified that studies with multiple behavioral outcomes have increased since 2007. To test this hypothesis, a dichotomous variable was created to indicate (yes/no) if the study reported multiple behavioral outcomes (e.g., substance use and sexual risk, or HIV testing and sexual risk). If multiple outcomes were found, study coders described them in a free-text field.

Assessment of Methodological Quality

Methodological quality was assessed according to CDC's Prevention Research Synthesis (PRS) criteria for evidence-based interventions (CDC, 2014c). The PRS team was created in 1996 to systematically review and summarize HIV behavioral intervention research. PRS aims to translate scientific evidence into evidence-based recommendations for HIV prevention (Higa et al., 2013; Lyles, Crepaz, Herbst, & Hay, 2006). PRS efficacy criteria were informed by efficacy criteria used in other projects such as the Community Guide and Grades of Recommendation, Assessment, Development, and Evaluation (GRADE). PRS prioritizes assessing the internal validity of trials "to ensure a reasonable level of confidence that the observed changes can be attributed to the intervention" (Higa et al., 2013; Lyles et al., 2006). Efficacy criteria can be applied to individual, group, couples, or community-level behavioral interventions and are comprised of five major domains: Intervention Description (e.g., clarity), Quality of Study Design (e.g., comparison arm, allocation), Quality of Study Implementation and Analysis (e.g., follow-up time, retention rate, alpha level, sample size), Strength of Evidence (e.g., statistically significant result [$p<.05$], relevant outcome, no harmful effects), and Additional Limitations to Evaluate (e.g., fatal flaws such as differential retention or substantial missing data). Studies that meet all criteria are classified as Evidence-Based Interventions either at the Good Evidence level or Best Evidence level. In 2013, Higa et al. of the PRS team reviewed MSM-specific interventions previously reviewed in the PRS database to better understand methodological challenges to demonstrating efficacy of HIV behavioral interventions for MSM. Higa et al. classified each trial into one of four mutually exclusive

categories: Evidence-Based Interventions (EBIs: trials that meet all efficacy criteria); Rigorous Non-EBIs (trials that meet all efficacy criteria except for a significant positive finding); Positive Non-EBIs (trials that reported a significant positive finding, but did not meet at least one other criterion); and Other Non-EBIs (trials that did not report significant positive finding and did not meet at least one other PRS criterion). This review uses original PRS criteria and Higa's additional four-level categorization system (Appendix C). Higa et al.'s ratings were entered into the data extraction form for trials prior to 2011. For trials not already rated by Higa, coders applied PRS criteria and resolved any disagreement by discussion.

Protocol Modifications

The original study protocol was modified during data extraction to make three important changes. The original scope of the review included any behavioral intervention trial with a comparison arm that targeted MSM with enough information to calculate an effect size. Upon reviewing the literature, three changes were made to the scope: 1) include only trials designed specifically for MSM (e.g., MSM representing at least 95 percent of the sample instead of trials that included >50 percent MSM), 2) exclude trials designed specifically for HIV-positive MSM (e.g. HIV-positive MSM comprising 100 percent of the sample), and 3) exclude any study that was not RCT (e.g., exclude quasi-experimental designs, even those with comparison groups). Changes were based on narrowing the scope of the review to best understand the effects of diverse interventions on MSM while minimizing threats to validity. Statistical challenges were found when attempting to isolate effects of the intervention on MSM if trials included non-MSM (e.g., reducing study weight to only

reflect the MSM subset), concerns emerged about increasing statistical and clinical heterogeneity if HIV-positive MSM were included, and concerns emerged about decreased methodological quality if non-RCT designs were included.

Statistical and Other Software

Microsoft Excel (2007), EndNote (X7), Biostat Comprehensive Meta-Analysis (Version 3), and SPSS (Version 22.0) were used to enter, manage, and analyze data for the systematic review, meta-analysis, and other data analyses. The data extraction protocol and standardized forms were created in Microsoft Excel. Descriptive statistics were calculated in Microsoft Excel and SPSS. Reference management was conducted exclusively in EndNote. Meta-analyses, moderator analyses, cumulative analyses, outlier analyses, and publication bias analyses were conducted in Biostat Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2015).

Statistical Analyses

Meta-Analysis

Meta-analysis was used as the primary analysis to synthesize results across trials and calculate a standardized effect size. One overall meta-analysis was conducted using the primary outcome from each study. Meta-regressions and subgroup analyses examined potential moderators of the overall effect size. Additional analyses included cumulative meta-analysis to calculate the cumulative effect size by year, homogeneity tests of the effect size distribution, analyses of publication bias, and sensitivity analysis for outlier analysis.

Outcome Measures

Outcomes used in this meta-analysis included: 1) mean number of episodes of UAI (*k*=14), 2) proportion of MSM reporting any UAI (*k*=18), 3) or mean number of male sexual partners for any anal intercourse (*k*=2). This study focused on unprotected anal intercourse (UAI) as the outcome of interest. UAI is the most epidemiologically important HIV risk behavior for MSM, and UAI data were available for most trials. Number of male sexual partners for anal intercourse was used only when the authors specified it as a primary outcome, and there was no available secondary outcome about UAI (*k*=2). This outcome is imperfect as it may measure partners for unprotected sex or protected sex, and may not be a reliable indicator of true HIV risk. Despite this limitation as an outcome measure, these studies were included because they contributed some meaningful measure of sexual risk.

Meta-analysis requires an assumption of independence between trials; only one outcome per trial is appropriate for calculating an overall effect size (Borenstein, Hedges, Higgins, & Rothstein., 2009). The primary outcome specified by trial authors was used to calculate one effect size for each trial. Fifteen studies reported multiple outcomes related to the same sexual risk construct, but did not specify a primary outcome. For these studies, the most precise outcome associated with highest risk of HIV acquisition was used. UAI confers the greatest risk for HIV and previous reviews have revealed count outcomes to be more sensitive than dichotomous outcomes (Johnson et al., 2008). Based on this rationale, *mean number of UAI episodes* was

specified as primary outcome. If count-level UAI data were not available, the *proportion of any UAI* was used.

Outcome variables used for meta-analysis did not discriminate between insertive or receptive anal intercourse (i.e., sexual position), primary or non-primary partners, or partner's HIV status (i.e., serodiscordant or seroconcordant partners). Specific results about behaviors with certain partners, or specific results about behaviors by sexual position (i.e., insertive vs. receptive anal intercourse) were used when more general results were unavailable. When UAI results were only presented by sexual position (e.g., insertive UAI vs. receptive UAI), results describing receptive anal intercourse were used because it is a higher risk behavior than insertive anal intercourse (k=2). When UAI results were only presented by primary vs. non-primary partner, non-primary partner results were used (k=1) under the assumption this situation conferred greater risk. Risk behavior outcomes specific to vaginal sex, oral sex, or sex with female partners were excluded from analysis. For trials that reported sexual behavior with both male and female partners as primary outcomes, secondary outcomes specific to anal intercourse with male partners were used (k=2).

Three trials tested more than two experimental conditions against one control group (Dilley et al., 2007; Hirschfield et al., 2012; Shoptaw et al., 2005). One approach to multiple comparisons is to split the comparison group into equal parts to compare each intervention to a control group. While this can preserve the assumption of independence, it can also create a bias towards homogeneity, especially if the study is large (Borenstein et al., 2009). While only one of the three studies was large, an alternative protocol was used. For the three studies, the most relevant intervention

condition was selected to compare to the control. Descriptive data for the excluded condition/s were not used in the meta-analysis and overall sample size for the study was adjusted.

For studies that reported outcomes at multiple time points, the last follow-up assessment was used to calculate effect sizes. Last follow-up was selected to examine sustained intervention effects. Most trials used a 3, 6, or 12 month follow up. Follow-up period was included as a moderator variable to statistically examine its relationship to the overall effect size.

Calculation of Effect Sizes

Effect sizes for each trial were calculated using the odds ratio (OR) for dichotomous measures and standardized mean difference (SMD) for count-level measures. Estimates of effect size were calculated based on descriptive data (e.g., means, standard deviations, frequencies, or proportions). When descriptive data were not available in any published report, requests for additional trial data were sent to trial authors. If data were not obtained from trial authors, the trial was excluded from analysis (k=4).

One measure of effect size is required to estimate the overall pooled effect size. To be consistent with previous meta-analyses using RCT designs, the OR was chosen as the common metric. SMDs were converted to the OR for the main meta-analysis. Disadvantages to using the OR for meta-analysis are that odds are less intuitive than risk, ORs are centered around 1, and ORs are problematic when cell size equals zero (Lipsey & Wilson, 2001). Advantages are that ORs have favorable mathematical properties, standard errors are easy to calculate, and they are commonly

reported in epidemiological research (Lipsey & Wilson, 2001; Borenstein et al., 2009; Higgins & Green, 2008). Calculations for odds ratios were conducted using the natural logarithm scale (lnOR). The lnOR, standard error (SE) of the lnOR, and confidence limits were calculated to generate values that are used in each step of the meta-analysis (Borenstein et al., 2009). Each lnOR was multiplied by its inverse variance weight, weighted lnORs across trials were summed, and then divided by the sum of the weights (Lipsey & Wilson, 2001). After meta-analysis, the results were converted back into OR units for display purposes.

Odds Ratios

For trials reporting dichotomous outcomes (e.g. frequencies or proportions), the odds ratio (OR) was calculated to estimate intervention effect. The OR was calculated for each cross-product ratio from 2x2 tables representing the frequency of UAI in the intervention group as compared to the control group at last follow-up (Herbst et al., 2007; Haddock, Rindskopf, & Shadish, 1988).

The computational formula for an odds ratio is (Borenstein et al., 2009):

$$OR = \frac{AD}{BC}$$

Each OR was then transformed to the log scale using the natural logarithm (lnOR) to obtain the log odds ratio (Johnson, 2002b; Bornstein et al., 2009):

$$LogOddsRatio = \ln(OddsRatio)$$

with approximate variance:

$$VLogOddsRatio = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

and approximate standard error:

$$SE_{LogOddsRatio} = \sqrt{V_{LogOddsRatio}}$$

As specified previously, the log odds ratio and its variance were used to generate the summary effect, confidence limits (*LL*=lower limits; *UL*=upper limits), and other measures in log units. Each value was then converted back to the odds ratio. Log odds ratios, variances, and 95% confidence intervals were transformed back to ORs by using the formula (Borenstein et al, 2009):

$$OddsRatio = \exp(LogOddsRatio),$$

$$LL_{OddsRatio} = \exp(LL_{LogOddsRatio}),$$

$$UL_{OddsRatio} = \exp(UL_{LogOddsRatio}).$$

Standardized Mean Differences

For trials with count-level outcomes (e.g., mean number of UAI events), standardized mean differences (SMD) and variances were calculated to estimate intervention effects (Johnson et al., 2002b; Bornstein et al., 2009). SMDs are the more appropriate option (as opposed to raw mean differences) when studies use different instruments to measure the outcome. Unadjusted, descriptive data (e.g., means, standard deviations, standard errors, frequencies, proportions) were used to calculate effect sizes. If measures of variance (i.e., standard deviation or standard error) were not reported, standard deviation was obtained by using sample sizes for intervention and control groups and the independent groups exact *p*-value. From these values, a *t*-value can be computed to obtain standard error which can be converted to standard deviation (Higgins & Green, 2008). The SMD (or Cohen's *d*) was the mean difference of intervention and control at follow-up divided by the within-group pooled standard deviation. Cohen's *d* is calculated by the formula: (Borenstein et al., 2009):

$$d = \frac{M_1 - M_2}{S_{within}}$$

The numerator values represent the sample means in each group and the denominator represents the pooled within-groups standard deviation calculated as:

$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

The variance of *d* is approximated by:

$$V_d = \frac{n_1 + n_2}{n_1} + \frac{d^2}{2\,(n_1 + n_2)}.$$

The standard error of *d* is the square root of V_d given as:

$$SE_d = \sqrt{V_d}.$$

SMD was converted to the OR to achieve a common metric, but has been shown to be upwardly biased when applied to small samples (Hedges, 1981). Hedges' *g* corrects for this bias (Lipsey & Wilson, 2001) with a correction factor (*J*), and provides an alternative effect size to Cohen's *d.* Hedges' *g* was additionally calculated by the following formula:

$$J = 1 - \frac{3}{4_{df} - 1},$$

$$g = J \times d,$$

$$V_g = J^2 \times V_d,$$

and

$$SE_g = \sqrt{V_g}.$$

To transform data, Cohen's *d* was converted to the logs odds ratio by standardized formulas (Johnson, 2002b):

$$\text{SMD}=\sqrt{3}\times[\ln(a)+\ln(d)-\ln(b)-\ln(c)]\div\pi,$$

$$\text{Var(SMD)}\approx 3/\pi^2\times\left[\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right],$$

$$\ln \text{OR}=\frac{\pi}{\sqrt{3}}\times SMD,$$

$$Var(lnOR)=\frac{\pi^2}{3}\times Var(SMD).$$

Intraclass Correlation Coefficient

For trials that used the community as the unit of assignment (i.e. community-level interventions, *k*=3), an estimate of the intraclass correlation coefficient (ICC) was applied to adjust study weights when cluster data was available. The ICC estimates the relative variability within and between clusters (Donner & Koval, 1980; Higgins & Greene, 2008). If the ICC is not accounted for, the variance of the intervention effect may be underestimated and the weight will be overestimated (Johnson et al., 2008). ICCs are not frequently available in study reports and it is common to use ICCs from other studies (Higgins & Green, 2008). To reduce each trial to the "effective sample size," an ICC of .005 was assumed because this value was cited in one community-level intervention report (Kelly et al., 1997; Johnson et al., 2002). To calculate the effective sample size, the original sample size is divided by the design effect:

$$Design\ Effect=1+(M-1)ICC,$$

where M is the average cluster size. Calculation of average cluster size was not possible for two of three trials due to missing cluster data and unadjusted results were used in the meta-analysis.

Model of Analysis

Variance was modeled using fixed and random effects models. Both models were specified *a priori* for comparison. Previous HIV prevention reviews have revealed highly homogeneous distributions so it was assumed that random and fixed models would yield identical results for the primary meta-analysis (Johnson et al., 2008.). Fixed effects models assume the variance is subject-level sampling error only (not study level variance) and are generally used for homogenous distributions. Random effects models are more conservative and assume that the variance has random study-level variance in addition to random subject-level sampling error. An additional random effects model was specified due to possible heterogeneity of trial subgroups, outcome measures, intervention content, and design variables (Hedges & Olkin, 1985). To fit a random effects model, non-iterative methods of the moment were used (Lipsey & Wilson, 2001). Biostat's Comprehensive Meta-Analysis was used to provide fixed and random effects mean effect sizes.

Calculations for the Mean Effect Size

A mean effect size, or summary effect size across studies, was calculated to obtain the most precise estimate of the mean using both fixed and random models. Before calculating the mean effect size, each effect size (lnOR) was weighted by the inverse of its variance to correct sample size bias. Effect sizes based on larger samples provide more precise estimates because the sampling error is smaller. Optimal weights are achieved by using the standard error (Hedges & Olkin, 1985). Individual studies were weighted by their respective sample sizes after all study effect sizes had been transformed to the lnOR. Each study was assigned a weight that was

the inverse of its variance, or the "inverse variance weight" (Lipsey & Wilson, 2001). The study weight is computed by this formula where V_{Yi} is the within-study variance for the study (i) Borenstein et al., 2009):

$$W_i = \frac{1}{V_{Y_i}}$$

For random effects analyses, study weights included within-study variance and an estimate of the between-studies variance (T^2).

The weighted mean effect size (M) was then calculated by multiplying each effect size by its weight, summing the products (W_iY_i), and then dividing by the sum of the weights (Borenstein et al., 2009):

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} w_i}$$

The variance of the mean effect size was calculated as the reciprocal of the sum of the weights (Borenstein et al., 2009):

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i},$$

and the standard error of the mean effect size was calculated by taking the square root of the variance (Borenstein et al., 2009):

$$SE_M = \sqrt{V_M}.$$

Confidence intervals (CIs) were set at 95% to examine the precision of the mean effect size and individual study effect size estimates. CIs demonstrate precision by providing the range of possible values for the effect size; a 95% confidence interval implies that the population mean effect size has a 95% probability of being in between the lower and upper bound of the CI. CIs were calculated by multiplying the standard

error by a critical z value to represent the proposed confidence level, adding the product to the point estimate for the upper limit, and subtracting the product from the point estimate for the lower limit. Upper *(UL)* and lower limits *(LL)* were given by these formulas (Borenstein et al., 2009):

$$LL_M = M - 1.96 \times SE_M \ ,$$

$$UL_M = M + 1.96 \times SE_M \ .$$

To test the null hypothesis that the mean effect size is zero, a *Z*-value was calculated (Borenstein et al., 2009):

$$Z = \frac{M}{SE_M},$$

and a *p*-value was given for the two-tailed test by (Borenstein et al., 2009):

$$p = 2\left[1 - \left(\Phi(|Z|)\right)\right],$$

where Φ(Z) represents the normal cumulative distribution.

Interpretation of Effect Sizes

Effect sizes are the primary result reported. After all calculations were conducted using the lnOR, values were converted back to the OR for display purposes. ORs are centered around 1 instead of zero with 1 indicating no relationship, or a null effect (Lipsey & Wilson, 2001). For behavioral trials in this study, ORs between 0 and 1 favor the intervention condition and indicate a positive effect. ORs greater than 1 favor the comparison group. ORs of 1 indicate no difference between intervention and comparison group. ORs are interpreted with their corresponding 95% CI to indicate the precision of the estimate. Wider confidence intervals indicate a less reliable estimate. Wider confidence intervals are associated with higher relative standard error and smaller samples. Confidence intervals that include the null value of

1 indicate there is insufficient evidence to detect a significant difference between groups. *P*-values were considered significant at the $p<.05$ level.

Cumulative Meta-Analysis by Year

Cumulative meta-analysis was performed to display how the weight of evidence has shifted over time. Cumulative meta-analyses are conducted by running a meta-analysis with one study, then repeating it with a second study added, and repeating again until all studies are included. Studies were sorted chronologically by year from oldest to newest, and overall percent change from earliest to latest year was calculated. Effect size change was also calculated from the earliest statistically significant cumulative effect size to the latest effect size. Forest plots were produced for visual examination of effect size with each new study added, and shifts in effect size over time. Cumulative meta-analysis is primarily a method for displaying results of a series of separate meta-analyses in one forest plot (Borenstein et al., 2009). To better understand how the evidence has accumulated over time, study year was also coded as a moderator and used in meta-regression to statistically evaluate the relationship between effect size and year.

Heterogeneity of Variance

Heterogeneity of variance implies that the effect size is not consistent across studies. Observed heterogeneity can occur because of random error (e.g. spurious or excess variation) or systematic differences (e.g. true variation) that require further analysis such as in the case of moderator variables. Variation between effect sizes was first examined by visually analyzing the forest plot to detect large differences between point estimates (OR) and the extent of the CI overlap. Large differences between ORs

or non-overlapping CIs suggest that random error is an unlikely explanation for the observed variation (Murad et al., 2014). Measures of Q and I^2 were used to statistically evaluate the true variance and proportion of real dispersion (Borenstein et al., 2009).

The Q statistic tests the homogeneity of the overall effect size with a null hypothesis that the underlying effect is the same across studies. Q is an approximate chi-square distribution with degrees of freedom equal to the number of studies minus one. Rejection of the null hypothesis (p<.05) suggests that the effect size distribution is heterogeneous and the variability is larger than what would be expected from random error; variability and most likely attributable to study characteristics or moderating variables. If the Q statistic is less than critical it suggests that the observed variance does not exceed what would be expected from sampling error alone.

However, there are limitations to the Q statistic. Power can be low with small samples and may miss the effects of study-level error (Lipsey & Wilson, 2001); in consideration of Q's limitations, P-values <.10 were considered significant to justify moderator analyses. Q is a standardized measure and was computed as (Borenstein et al., 2009):

$$Q = \sum_{i-I}^{k} W_i(Y_i - M)^2,$$

where W_i is the study weight, Y_i is the study effect size, and M is the overall effect size.

The second statistical test used I^2 where:

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%.$$

I^2 is a measure of the magnitude of variability as opposed to a measure of statistical significance like the *Q* statistic (Murad et al., 2014). I^2 is a measure of the real dispersion due to heterogeneity over random error alone (Higgins & Green, 2008), or the ratio of excess dispersion to total dispersion (Borenstein et al., 2009). When I^2 is close to 0%, it is likely that observed variability between point estimates is due to random error. The further I^2 moves from 0, the more likely that variability is explained by systematic differences requiring subgroup or moderator analyses. Interpretation of I^2 needs to consider not only magnitude and direction of effects, but also the *Q* statistic and CIs for I^2. Guidelines to interpret I^2 vary, but rough guidelines were used (Higgins et al., 2003):

1. 25%=low heterogeneity
2. 50%= moderate heterogeneity
3. 75% =high heterogeneity.

Statistical Tests of Moderators

Four design variables were specified as moderators *a priori* based on previous work from Johnson et al., 2008*:* 1) outcome measure (count-level vs. dichotomous), 2) comparison condition (non HIV-related vs. HIV-related vs. waitlist), 3) and intervention time span (≤1 month vs. >1 month) and retention rate (better in comparison vs. equal or intervention). Hypotheses specified that count-level measures, non-HIV-related controls, interventions with shorter time spans, and interventions with better retention in the treatment group would yield stronger effects across all studies. In addition to *a priori* variables, participant characteristics and study characteristics were analyzed as potential moderators for exploratory purposes.

Variables tested for moderator analysis were categorical and continuous variables and required different statistical tests. Categorical variables were tested using the analog to analysis of variance (ANOVA). Continuous variables were tested using weighted meta-regression analysis. Subgroup analyses used mixed-effects models which use a fixed effect model across subgroups and a random effects model within subgroup (Borenstein et al., 2009). The analog to ANOVA separates the total homogeneity (*Q)* into between-group variance and within-group variance. Subgroup analyses were computed by the general analog ANOVA given by:

$$Q_w = \sum_{i=1}^{k} w_i(ES_i - M_j)^2$$

where Q_w is the pooled group (within) variance and Q_B is the between groups variance:

$$Q_B = \sum_{i=1}^{k} w_j M_j^2 - \frac{(\Sigma_j M_j)^2}{\Sigma w_j}.$$

w_j is equal to the sum of weights for each subgroup, M_j refers to the weighted mean effect size for each subgroup, w_i is equal to the study weight ($\frac{1}{SE^2}$), ES_i refers to the study effect size, *M* is the summary effect, *k* is the number of studies (effect sizes) and *j* refers to the number of groups (Borenstein et al., 2009; Lipsey & Wilson, 2001).

Continuous moderators were tested using the method of moments for meta-regression analyses. Meta-regression assesses heterogeneity by developing regression models with independent variables that represent individual study characteristics and the dependent variable is the effect size (Borenstein et al., 2009; Murad et al., 2014). Comparable to the analog ANOVA, variables were first assessed by a *Q* test; *Q* was separated into variability accounted for by the: 1) regression equation (Q_R), and 2) the

residual variance (Q_B) (Borenstein, 2009). True variance was explained by R^2 given by:

$$R^2 = 1 - \left(\frac{T^2_{unexplained}}{T^2_{total}}\right)$$

where R^2 is the variance of true effect sizes across studies and T^2 is Tau-squared. (Borenstein et al., 2009).

Outliers and Sensitivity Analysis

Outlier analysis was performed to identify any study effect sizes that were extreme and could distort the mean effect size or variance. The forest plot was visually analyzed to identify extreme observations or very large sample sizes. Potential outliers were then checked for data entry errors by performing a secondary review of the manuscript and a second data extraction. Outliers found to be three standard deviations from the mean (z=>3.0) were tested using sensitivity analysis where the meta-analysis was run with outliers removed and compared to the original meta-analysis. Sensitivity analysis was also performed using a "one study removed" analysis where the meta-analysis was run with all studies except the first, then all studies except the second, and then the third, until all studies were run (Borenstein et al., 2009). Forest plots were produced to plot effect sizes as each study was added and their 95% CIs. Plots were visually examined to detect any shift in the effect sizes. Studies found to be problematic were excluded and the meta-analysis was re-run to compare results.

Publication Bias

Publication bias addresses the larger research issue that not all completed trials are published. Trials accepted for publications are more likely to report large

treatment effects than studies reporting modest or null effects (Dickersin & Min, 1993; Borenstein et al., 2009). This publication bias can overestimate the treatment effect. Every meta-analysis is vulnerable to publication bias and requires a combination of methods to detect the impact of bias on the overall effect size. This review included only peer-reviewed published articles as a method of quality control to ensure review of the highest quality evidence. Therefore, this review may have a high risk of the "file drawer" problem (Rosenthal, 1979) where the effect size is overestimated due to studies with lower effect sizes being excluded. However, publication bias is thought to be less likely in fields of public health importance such as HIV prevention because results with small samples, small effects, or insignificant results are often published (Marsh et al., 2001).

To test for publication bias and study its impact, six tests were conducted: 1) the funnel plot, 2) Egger's regression index, 3) Duval and Tweedie's trim and fill test, 4) Rosenthal's *fail-safe* N, 5) Orwin's adapted version of Rosenthal's *fail-safe* N, and 6) cumulative meta-analysis based on trial standard error. Additionally, trials were coded as yes/no to indicate statistically significant positive effects.

Funnel Plot

The funnel plot was examined first. The funnel plot is a scatterplot that plots the intervention effect size against the standard error or other measure of study precision (i.e. sample size). Funnel plots allow for a quick visual display of symmetry. Effect sizes were plotted on the X-axis and standard errors were plotted on a reverse scale on the Y-axis. In the absence of publication bias, the scatter plot looks like an inverted, symmetrical funnel. Typically, large studies (or studies with smaller

standard errors) cluster towards the top of the graph and around the mean effect size (Borenstein et al., 2009; Higgins & Green, 2008). Smaller studies appear towards the bottom part of the graph and are often widely dispersed and fall to the right and left of the mean. When an asymmetrical pattern appears with gaps on the left or right of the mean, it suggests the presence of publication bias. The asymmetrical shape can suggest that smaller studies were more likely to be included if they have larger than average effects. These studies most likely met criterion for statistical significance and were favored in publication over smaller studies with moderate results.

Visual interpretation of symmetry can help detect a relationship between sample size and effect size, but it does not explain the relationship (Gleser & Olkin, 1996; Duval & Tweedie, 2000; Pham et al., 2000). For example, the studies may represent a biased sample of smaller studies, or it is possible that the observed larger effect size is truly larger in these studies due to study quality. Additional tests were conducted to study "small study effects" and examine if small-study effects are caused by methodological quality, true heterogeneity, study artifacts, or chance (Higgins & Green, 2008; Egger 1997a).

Egger's Test of the Intercept

Egger's test of the intercept was conducted to yield a quantitative estimate of the amount of bias observed in the funnel plot. Egger's test is calculated by regressing the standardized effect (i.e., the effect size divided by standard error) on the inverse of the standard error (i.e., precision estimate). For this equation, the size of the treatment effect is captured by the slope of the regression line (B1) and bias is captured by the intercept (B0) (Borenstein et al., 2009). Studies with higher standard errors produce

precision estimates closer to zero whereas studies with lower estimates of standard error produce precision estimates further from zero, suggesting studies with more robust sample sizes. Bias is detected by a regression line that does not approach the intercept of origin.

Duval and Tweedie's Trim and Fill

Duval and Tweedie's Trim and Fill method was used to impute studies that were indicated as missing from the funnel plot (Duval & Tweedie, 2000). This analysis was used to determine where the studies were missing on the plot, impute their estimated values to the analysis, and then re-compute the combined effect (Borenstein et al, 2009). The trim and fill method is an iterative procedure to remove, or trim, extremely small studies from the right side of the plot while re-computing the overall effect size with each iteration. Iterations continue until a symmetric funnel plot is achieved for the revised effect size. Trimming reduces variance of effects and shrinks confidence intervals. To compensate, the original studies are added (i.e., filled) back into the analysis, imputing a mirror image for each. A funnel plot was produced that included both the original and imputed studies to visually examine any effect size shift when imputed studies are included. The Trim and Fill method is very sensitive to outlier studies, and has strong assumptions about reasons for the missing data.

Rosenthal's fail-safe N and Orwin's adapted version of the fail-safe N

Rosenthal's *fail-safe* N and Orwin's adapted version of the *fail-safe* N were conducted to assess the magnitude of the effect of publication bias and account for the file-drawer problem (Rosenthal, 1979; Begg & Mazumdar, 1994). Rosenthal's

method computes the number of studies required to nullify the effect. When this number is relatively small, publication bias is a potential threat. If this number if large, it suggests the intervention effect is not nil. Rosenthal's method needs to be interpreted with caution because it assumes that the effect in the missing studies is nil, as opposed to the effect of these studies being in the opposite direction, therefore the true number of studies required to nullify the effect may be smaller than the calculated *fail-safe* N. Additionally, the *fail-safe* N is a test based on statistical significance and does not account for clinical significance. This study calculated the fail-safe N using the modern practice of computing a summary effect, combining the effect sizes, and then computing a p-value for the combined effect sizes. To address limitations of the *fail-safe* N, Orwin's fail-safe N was calculated and treated as a more reliable measure of the number of studies needed to nullify the effect (Orwin, 1983). Orwin's fail-safe N differs from Rosenthal's because it determines the number of studies using a predetermined effect size set by the researcher, rather than an arbitrary *p*-value. This allows the researcher to estimate how many missing studies are needed to bring the overall effect size down before the overall effect would be trivial. In this study, the criterion odds ratio that defined a trivial effect was set as 1.0 and the missing study mean odds ratio was assumed to be 1.5.

Cumulative Meta-Analysis by Standard Error

Publication bias is assumed to be related to smaller studies or small-study effects. The effect of study size was also examined using a cumulative meta-analysis. This procedure is generally used to display shifts in the cumulative weight of the evidence over time, but may also be used to further investigate impact of

publication bias or small-study effects (Borenstein, Higgins, Rothstein, & Hedges, 2015). To use this procedure for publication bias, trials are sorted by precision, or their standard error, in order of the most precise to least precise (i.e. smallest standard errors to largest standard errors). Larger standard errors roughly correspond to smaller sample size and less precision, so if the odds ratio does not shift with the addition of the smaller studies, then there is no evidence to suggest publication bias related to small-study effects. After trials were added from smallest to largest standard error, cumulative meta-analysis was performed as each new study was added. Forest plots were produced to display cumulative effect sizes with each new study. Visual examination was performed to detect any shift in overall effect size when the smaller trials (i.e. larger standard errors or less precise) were added. There is valid concern for bias if the point estimate shifts with the addition of smaller trials, or trials with less precision.

Missing data

For this study, missing data was related to reporting bias and refers to data needed to calculate effect sizes for positive or negative effects in studies otherwise eligible for review. Missing data can also refer missing values specific to covariates used for subgroup analyses or meta-regressions. Effect sizes were calculated from descriptive statistics when possible (e.g. sample size for intervention and comparison groups, means, standard deviations, proportions). When descriptive statistics were not available, values were converted from standard errors, *t, F,* exact *p*-values, or 95% confidence intervals as appropriate (Higgins & Green, 2008). When data were not available for effect size calculation, requests for descriptive data were sent to study

authors. When study authors did not respond, requests were sent to authors of previous meta-analyses who may have previously retrieved missing data from study authors (e.g. CDC's Prevention Research Synthesis Team). If study authors did not respond, and effect sizes could not be accurately calculated, studies were excluded from meta-analysis rather than performing imputation. Case analysis was used for moderator data. Data were not imputed for moderator analyses, and studies with missing data were excluded.

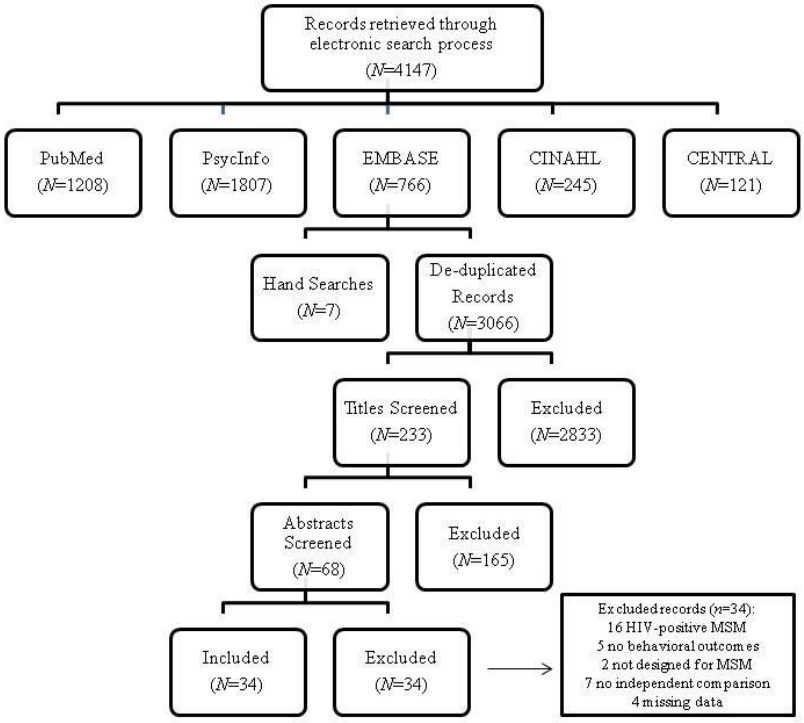
CHAPTER 5: RESULTS

Search Results

The search process is summarized in Figure 1. The search retrieved 4,147 records from electronic databases as potentially relevant citations: PubMed (1,208), PsycINFO (1,807), EMBASE (766), CINAHL (245), and CENTRAL (121). De-duplication resulted in 3,059 records. Seven additional records were found through hand searches resulting in 3,066 records for review. Manual review of record titles resulted in 233 records deemed to be potentially relevant. Manual review of record abstracts resulted in 108 records potentially eligible for inclusion and these manuscripts were retrieved. After manual review of full text, 40 records were excluded because they did not meet eligibility criteria. Sixty-eight manuscripts were retained for full review by two coders.

During coding, the protocol was revised to add additional exclusions to eligibility. Full review with new protocol exclusions resulted in thirty manuscripts excluded due to having 100% HIV-positive MSM (k=16), not reporting a behavioral outcome (k=5), not being specifically designed for MSM (k=2), or not having an independent comparison group (k=7). A total of 38 records were selected for inclusion and effect size data were examined. Eight records were identified as eligible for inclusion, but missing essential data. All eight authors were contacted and missing data were requested. Three authors responded and provided data to calculate an effect size for the specified primary outcome, and a previous meta-analysis supplied effect size data for one trial. Four trials were excluded due to missing data resulting in a final sample size of 34 trials (Table 1)

Figure 1
Search and Retrieval Process Results



Description of Trials

An overview of trials are summarized in Table 1. Descriptive characteristics are summarized in Table 2. A total of 17,872 participants were enrolled at baseline across 34 trials (median sample size=328, range=50-4,295). Across all trials, a median of 66 percent of participants were white (*M*=50.8, *SD*=35.8). A median of nine percent of participants were black (*M*=29.5, *SD*=39.0), and 11 percent were Latino (*M*=17.0, *SD*=25.3). Mean age was 33.3 (*SD*=5.4, median=34) years with a mean range of 21.3-42.8 years. A median of 31 percent were high school graduates or less, and a median of 15 percent were HIV-positive. Across 34 trials, only nine (26.4 percent) focused specifically on minority racial or ethnic groups (6=black,

2=Hispanic/Latino, 1=Asian), four (11.8 percent) focused specifically on young adult MSM, and 6 (17.6 percent) focused specifically on substance-using MSM. No studies specifically focused on male sex workers, and only one study focused on men who have sex with men and women (MSMW). A median of 74 percent of participants reported any substance use and 42 percent reported mental health distress or history of victimization. Most participants self-identified as gay or bisexual (median=98 percent). The majority of trials were conducted in the Western region of the U.S. (*k*=16), 11 trials were conducted in the Northeast, five trials were conducted in the Midwest, five were conducted in the South, and four recruited national samples using the internet or telephone. Four trials had multiple trial sites in more than one U.S. region. The most frequently reported trial sites were in cities with higher proportions of MSM such as New York City, San Francisco, Los Angeles, and Seattle.

Eighteen trials (52.9 percent) contributed count-level primary outcomes and 16 (47.0 percent) contributed dichotomous outcomes as primary outcomes. Most trials were either group-level interventions (*k*=17) or individual-level interventions (*k*=14). Three interventions were community-level interventions. Median intervention time span was 3 weeks (range 1-288 weeks) with a median of 4 sessions (range:1-48) and median of 8 hours (range: 1-96). The majority of trials used an HIV-related comparison group (*k*=17), seven used a non-HIV related comparison group, and 10 used a wait-list comparison group. More trials reported ≥80 percent overall retention rates (*k*=19) than <80 percent retention (*k*=14). One study did not clearly report retention rates and was excluded from retention analysis. Of trials that reported retention rates by experimental condition (*k*=30), most trials (*k*=19 or 55.8 percent)

reported equal or about equal retention rates, 10 trials reported better retention in the comparison group, and no trials reported better retention in the intervention group.

Intervention content varied across trials. The majority of trials reported intervention content that included the development of individualized risk reduction plans (*k*=27). About 68 percent of trials (*k*=23) reported intervention content that focused on increasing sexual communication, 58.8 percent (*k*=20) reported content focused on increasing skills, and 41.1 percent (*k*=14) reported intervention content focused on the impact of stigma or discrimination. Nine trials reported having multiple behavioral outcomes of interest other than sexual risk reduction. Six focused on additional outcomes of substance use reduction and three focused on increasing HIV testing. Of these trials, six or 66.6 percent were conducted after the last review in 2007.

About half of trials (*k*=21, 62 percent) reported statistically significantly different findings ($P<.05$) between experimental conditions on at least one relevant outcome at last follow-up assessment. Eleven of 34 trials were found to be previously classified as Evidence-Based Interventions (EBI) by CDC PRS and listed in the Compendium (CDC, 2014b). Of the remaining 23 trials, 10 were classified as Positive Non-EBIs, seven were Rigorous Non-EBIs, and six were Other Non-EBIs.

Table 1
Overview of Included Trials

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure*	Control Group	Follow-Up	Quality Rating
Carballo-Dieguez et al.	2005	180 Latino MSM in New York City	8 weekly two-hour sessions	GLI	No. UAI events*, % UAI	Wait-list	12 months	Other Non-EBI
Carpenter et al.	2010	112 MSM aged 18-39 recruited online	Web-based skills training, 1 two-hour session	ILI	No. UAI events	Non-HIV related	3 months	Positive Non-EBI
Choi et al.	1995	329 Asian Pacific Islander MSM in San Francisco	1 three-hour safer sex skills training group	GLI	% UAI*, No. male partners	Wait-list	3 months	EBI
Coffin et al.	2014	326 substance-using MSM in San Francisco	1 one-hour session of PCC	ILI	No. UAI events*, No. male partners	HIV-related	6 months	Rigorous Non-EBI
Dilley et al.	2002	305 HIV-negative MSM in San Francisco	1 one-hour session of PCC with paraprofessional counselor	ILI	No. UAI events	HIV-related	12 months	EBI
Dilley et al.	2007	248 HIV-negative MSM in San Francisco	1 one-hour session of PCC, sexual diary	ILI	% UAI	HIV-related	12 months	Rigorous Non-EBI
Eaton et al.	2011	149 HIV-negative MSM in Atlanta	1 one-hour session with peer counselors to address serosorting risks	ILI	No. male partners	HIV-related	3 months	EBI

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure	Control Group	Follow-Up	Quality Rating
Harawa et al.	2013	437 black MSMW in Los Angeles	MAALES 6 two-hour small group sessions	GLI	No. UAI events	HIV-related	6 months	EBI
Hightow-Weidman et al.	2012	50 young black MSM in North Carolina	Four 30-minute weekly sessions by website	ILI	No. male partners*, % UAI	HIV-related	3 months	Other Non-EBI
Hirshfield et al.	2012	3092 MSM recruited online	brief HIV prevention video or webpage	ILI	% UAI events	HIV-related	2 months	Positive Non-EBI
Kegeles et al.	1996	300 young MSM in Oregon	Peer-led outreach, small groups, and media campaign	CLI	% UAI events	Wait-list	12 months	EBI
Kelly et al.	1989	104 HIV-negative MSM in mid-sized city	Twelve 90-minute weekly CBT group sessions	GLI	No. UAI events*, % condom use, No. partners	Wait-list	4 months	Positive Non-EBI
Kelly et al.	1991	659 MSM in Mississippi and Louisiana	POL peer-led community intervention	CLI	% UAI events*, % condom use, No. male partners	Non-HIV related	12 months	EBI
Kelly et al.	1997	442 MSM in 8 US cities	POL peer-led community intervention	CLI	No. UAI events*, % UAI; No. male partners	HIV-related	12 months	EBI
Koblin et al.	2012	283 black MSM in New York City	5 two-hour risk reduction sessions with meal prep	GLI	% UAI events	Non HIV-related	3 months	Rigorous Non-EBI

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure	Control Group	Follow-Up	Quality Rating
EXPLORE	2004	4295 MSM in six US cities	10 counseling sessions with 3-month booster session for 4 years	ILI	% UAI events	HIV-related	48 months	EBI
Kurtz et al.	2013	515 substance-using MSM in South Florida	4 two-hour small group weekly sessions about substance use and sexual risk	GLI	No. UAI events*, No. male partners	HIV-related	12 months	Rigorous Non-EBI
Lui et al.	2013	400 HIV-negative MSM in San Francisco, Atlanta, and Boston	PrEP, HIV testing, risk reduction counseling every 3 months	ILI	% UAI events	Wait-list	24 months	Other Non-EBI
Mansergh et al.	2010	1686 substance using MSM in 4 US cities	6 two-hour CBT group sessions focused on substance use and sexual risk	GLI	% UAI events	Non HIV-related	12 months	Rigorous Non-EBI
Menza et al.	2010	127 MSM who use methamphetamine in Seattle	12 weeks of bi-weekly contingency management	ILI	% UAI events*, No. male partners	Non HIV-related	6 months	Rigorous Non-EBI
Mustanski et al.	2013	102 young MSM in Chicago	3 two-hour sessions of KIU! online program focused on HIV testing and sexual risk	ILI	No. UAI events	HIV-related	3 months	Positive Non-EBI

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure	Control Group	Follow-Up	Quality Rating
O'Donnell et al.	2014	370 Latino MSM in New York City	Adaptation of VOICES/VOCES (1 hour video) into No Excuses/Sin Buscar Excuses	GLI	No. UAI events*, % condom use	HIV-related	3 months	EBI
Parsons et al.	2014	143 substance using young MSM in New York City	4 sessions of MI delivered over 12 weeks focused on substance use and sexual risk reduction	LI	No. UAI events	HIV-related	12 months	EBI
Peterson et al.	1996	318 HIV-negative African-American MSM in San Francisco	1 group or 3 three-hour weekly small groups using CBT self-management training	GLI	% UAI events	Wait-list	18 months	Positive Non-EBI
Picciano et al.	2001	89 MSM in Seattle	Single session 90-minute telephone-based brief counseling using MI	ILI	No. UAI events*, No. male partners	Wait-list	2 months	Rigorous Non-EBI
Picciano et al.	2007	319 MSM in Seattle and Portland	Three 90-minute sessions of MET delivered by telephone over 6 weeks	ILI	No. male sex partners	HIV-related	10 months	Other Non-EBI

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure	Control Group	Follow-Up	Quality Rating
Roffman et al.	1997	548 MSM recruited by telephone	14 weekly 90-minute group CBT focused on relapse prevention by telephone	GLI	% UAI events*, No. UAI events, No. male partners	Wait-list	3 months	Positive Non-EBI
Roffman et al.	1998	159 MSM in Seattle	17weekly sessions of group CBT counseling focused on relapse prevention	GLI	No. UAI events*, No. male partners	Wait-list	1 months	Positive Non-EBI
Simon Rosser et al.	2002	422 MSM in a Midwestern city	2-day comprehensive human sexuality seminar	GLI	% UAI events	HIV-related	12 months	Positive Non-EBI
Shoptaw et al.	2005	162 methamphetamine dependent MSM in Los Angeles	16 weeks of tri-weekly 90-minute culturally tailored group CBT	GLI	No. UAI events*, % UAI	Non HIV-related	12 months	Other Non-EBI
Stall et al.	1998	456 MSM in substance use treatment in San Francisco	16 unstructured weekly 3-hour treatment groups focused on coping skills	GLI	% UAI events	Non HIV-related	15 months	Other Non-EBI
Tobin et al.	2013	188 African-American MSM in Baltimore	6 bi-weekly 2-hour culturally tailored group modules plus one individual session	GLI	% UAI events	HIV-related	3 months	Positive Non-EBI

Publication	Year	Participants at Baseline	Intervention Description	Level	Effect Measure	Control Group	Follow-Up	Quality Rating
Valdiserri et al.	1989	584 MSM in Pittsburgh	Small group lecture plus skills training for safer sex negotiation	GLI	No. male sex partners	HIV-related	12 months	Positive Non-EBI
Wilton et al.	2009	338 HIV-negative black MSM in New York City	3MV 6 three-hour sessions delivered over 3 days as a weekend retreat	GLI	No. UAI events*, No. male partners	Wait-list	6 months	EBI

Notes: *effect measure extracted for meta-analysis based on primary outcome or meta-analysis protocol
No.=number
%=percent
UAI=unprotected anal intercourse
MSM=men who have sex with men
MSMW=men who have sex with men and women
ILI=individual-level intervention
GLI=group-level intervention
CLI=community level intervention
EBI=evidence-based intervention
PCC=personalized cognitive counseling
MAALES=Men of African American Legacy Empowering Self
CBT=cognitive behavioral therapy
POL=Popular Opinion Leader
PrEP=pre-exposure prophylaxis
KIU!=Keep It Up!
MI=motivational interviewing
MET=motivational enhancement therapy
3MV=Many Men Many Voices.

Table 2
General Characteristics of Trials

	k (34)	Percent	Mean	*SD*
Conducted Pre or Post ART				
Pre-ART	11	32.3		
Post-ART	23	67.6		
MSM Subgroup				
Young MSM	4	11.8		
Racial/ethnic minority MSM	9	26.5		
Substance-using MSM	6	17.6		
Other or none specified	15	44.1		
Region				
Northeast	11	32.3		
South	5	14.7		
Midwest	5	14.7		
West	16	47.0		
National	4	6.8		
Sample Characteristics				
Median % black	29	9	29.5	39.0
Median % Latino	29	11	17.0	25.3
Median % white	34	66	50.8	35.8
Median % gay or bisexual	20	98	94.5	7.6
Median % HIV-positive	30	15	19.5	22.0
Median % high school or less	25	31	35.4	19.6
Median % mental health	5	42	56.5	25.8
Median age	34	34	33.3	5.4
Intervention Level				
Individual	14	41.1		
Group	17	50.0		
Community	3	8.8		
Comparison Group				
Wait-list	10	29.4		
HIV-related	17	50.0		
Non HIV-related	7	20.6		
Outcomes				
Count-level*	18	52.9		
Dichotomous*	16	47.0		
Multiple outcomes reported	9	26.5		
P<.05 result reported	21	61.7		
Retention				
>80% overall retention	19	55.9		
Better in comparison condition	11	32.3		
Better in intervention condition	0	0		
About equal	19	55.9		
Follow-Up Assessment				
<3 months	12	35.2		

Table 2
General Characteristics of Trials

	k (34)	Percent	Mean	*SD*
≥3 months	22	64.7		
Intervention Dose				
Median time span in weeks	33	3	14.9	49.6
Median total time in hours	28	8	12.9	21.3
Median # of sessions	33	4	7.3	10.1
Intervention Characteristics				
Pilot tested	19	55.8		
Conducted in MSM setting	6	17.6		
Delivered by peers	6	17.6		
Focus on sexual communication	23	67.6		
Focus on skills building	18	52.9		
Focus on individual plans	27	79.4		
Focus on stigma	13	38.2		
Use of technology	8	23.5		
Evidence Level				
Evidence-Based Intervention	11	32.3		
Positive Non-EBI	10	29.4		
Rigorous Non-EBI	7	20.6		
Other Non-EBI	6	17.6		

Notes: * outcome used in meta-analysis as primary outcome

Overall Meta-Analysis

Meta-analysis was conducted using data extracted from all 34 studies. Count-level outcomes were used in 18 (52.9 percent) trials and dichotomous outcomes were used in 16 (47.1 percent) trials. Effect sizes resulted from a total of 13,272 participants. Figure 2 displays the descriptive data, ORs, confidence intervals, and outcome measures from each trial. Trials were sorted in descending order by year. Effect sizes ranged from *OR*=0.431 to *OR*=1.343. Under the random effects model the mean effect size was *OR* =0.859 (95% CI [0.790, 0.933]). This result was significantly different than zero ($p<.001$). The fixed effects model yielded a statistically significant result consistent with random effects (*OR*=.865, 95% CI [0.815, 0.919], $p<.001$). Test of the *Q* statistic indicated little heterogeneity and was

not significant (Q[33]=39.35, p =.207). I^2 was low indicating relatively little heterogeneity (I^2=16.14).

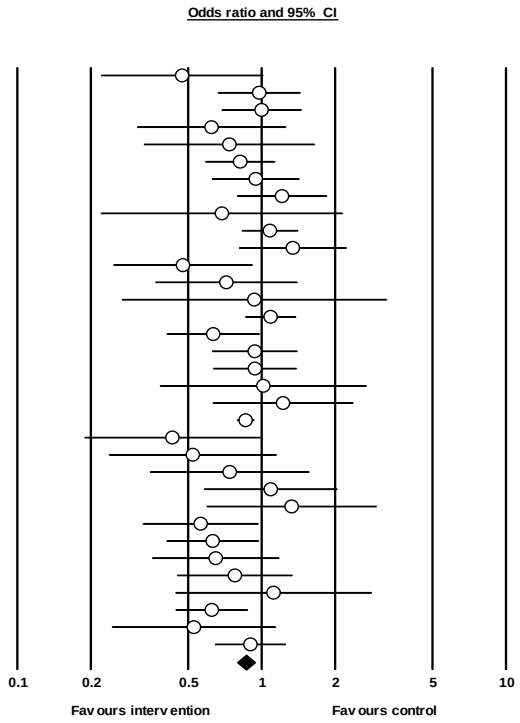
Cumulative Meta-Analysis

Cumulative meta-analysis was conducted on all 34 studies sorting by publication year from earliest to latest (Figure 3). Effect size shrinkage was determined by calculating the percent change from the first cumulative effect size that became significant (third row: Kelly, 1991) to the last row. A gradual effect size shrinkage of 19.5 percent was observed from 1991 (*OR*=.719) to 2014 (*OR*=.859). The forest plot reveals the largest effect size shift to occur after the inclusion of EXPLORE in 2004 from .720 (95% CI [0.619, 0.838]) to .777 (95% CI [0.688, 0.878]) resulting in a relative change over 1.0. After EXPLORE, the cumulative effect size stabilized and confidence intervals narrowed. The effect size moved closer to the null value of 1.0 with the addition of each new trial. Results warrant statistical analysis of year as a covariate in meta-regression analyses.

Figure 2
Overall Meta-Analysis Results

Effect Sizes of Behavioral Interventions for US MSM, 1988-2014

Study name	Outcome	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value
Parsons, 2014	% UAI events	0.473	0.219	1.021	-1.908	0.056
Coffin, 2014	# UAI events	0.978	0.659	1.449	-0.113	0.910
Odonnell, 2014	# UAI events	1.000	0.682	1.466	0.000	1.000
Mutanski, 2013	# UAI events	0.624	0.308	1.265	-1.307	0.191
Tobin, 2013	% UAI events	0.738	0.328	1.660	-0.734	0.463
Kurtz, 2013	# UAI events	0.817	0.585	1.141	-1.186	0.236
Harawa, 2012	# UAI events	0.946	0.623	1.435	-0.262	0.794
Lui, 2013	% UAI events	1.212	0.789	1.861	0.880	0.379
Hightow-Weidman, 2012	# sex partners	0.688	0.219	2.161	-0.641	0.521
Hirshfield, 2012	% UAI events	1.082	0.825	1.418	0.570	0.569
Koblin, 2012	% UAI events	1.343	0.806	2.239	1.131	0.258
Eaton, 2011	# sex partners	0.477	0.246	0.924	-2.195	0.028
Carpenter, 2010	# UAI events	0.717	0.365	1.408	-0.966	0.334
Menza, 2010	% UAI events	0.933	0.266	3.269	-0.108	0.914
Mansergh, 2010	% UAI events	1.089	0.853	1.391	0.683	0.494
Wilton, 2009	# UAI events	0.634	0.407	0.987	-2.019	0.043
Dilley, 2007	# UAI events	0.937	0.624	1.409	-0.311	0.756
Picciano, 2007	# sex partners	0.938	0.630	1.397	-0.313	0.754
Shoptaw, 2005	# sex partners	1.016	0.382	2.702	0.031	0.975
Carballo-Dieguez, 2005	# UAI events	1.223	0.628	2.382	0.592	0.554
EXPLORE, 2004	% UAI events	0.860	0.788	0.938	-3.401	0.001
Dilley, 2002	% UAI events	0.431	0.187	0.993	-1.975	0.048
Simon Rosser, 2002	% UAI events	0.523	0.236	1.158	-1.597	0.110
Picciano, 2001	# UAI events	0.740	0.348	1.576	-0.780	0.435
Roffman, 1998	# UAI events	1.089	0.578	2.052	0.265	0.791
Stall, 1998	% UAI events	1.327	0.593	2.973	0.688	0.491
Roffman, 1997	% UAI events	0.563	0.325	0.976	-2.045	0.041
Kelly, 1997	# UAI events	0.630	0.406	0.979	-2.056	0.040
Kegeles, 1996	% UAI events	0.649	0.355	1.188	-1.402	0.161
Choi, 1996	% UAI events	0.777	0.449	1.347	-0.899	0.369
Peterson, 1996	% UAI events	1.119	0.442	2.832	0.237	0.813
Kelly, 1991	% UAI events	0.625	0.443	0.883	-2.668	0.008
Kelly, 1989	# UAI events	0.528	0.243	1.149	-1.610	0.108
Valdiserri, 1989	# sex partners	0.900	0.641	1.263	-0.610	0.542
		0.859	0.790	0.933	-3.580	0.000

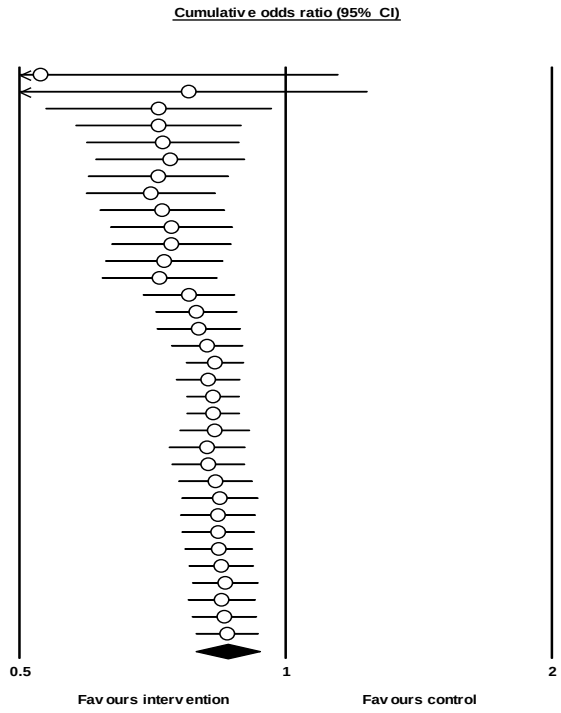


k=34, random effects model

Figure 3
Cumulative Meta-Analysis Results

Cumulative Effect Sizes by Trial Year of Publication, 1988-2014

Study name	Outcome	Cumulative statistics				
		Point	Lower limit	Upper limit	Z-Value	p-Value
Kelly, 1989	# UAI events	0.528	0.243	1.149	-1.610	0.108
Valdiserri, 1989	# sex partners	0.777	0.487	1.239	-1.059	0.289
Kelly, 1991	% UAI events	0.719	0.534	0.966	-2.188	0.029
Kegeles, 1996	% UAI events	0.718	0.578	0.893	-2.981	0.003
Choi, 1996	% UAI events	0.726	0.594	0.888	-3.123	0.002
Peterson, 1996	% UAI events	0.740	0.608	0.901	-3.002	0.003
Roffman, 1997	% UAI events	0.718	0.597	0.864	-3.514	0.000
Kelly, 1997	# UAI events	0.704	0.594	0.835	-4.036	0.000
Roffman, 1998	# UAI events	0.725	0.615	0.855	-3.829	0.000
Stall, 1998	% UAI events	0.743	0.632	0.873	-3.614	0.000
Picciano, 2001	# UAI events	0.743	0.634	0.870	-3.697	0.000
Dilley, 2002	% UAI events	0.729	0.624	0.851	-3.999	0.000
Simon Rosser, 2002	% UAI events	0.720	0.619	0.838	-4.231	0.000
EXPLORE, 2004	% UAI events	0.777	0.688	0.878	-4.062	0.000
Shoptaw, 2005	# sex partners	0.793	0.711	0.883	-4.218	0.000
Carballo-Dieguez, 2005	# UAI events	0.797	0.714	0.891	-3.998	0.000
Dilley, 2007	# UAI events	0.815	0.741	0.897	-4.207	0.000
Picciano, 2007	# sex partners	0.832	0.770	0.899	-4.679	0.000
Wilton, 2009	# UAI events	0.817	0.750	0.890	-4.635	0.000
Carpenter, 2010	# UAI events	0.828	0.771	0.889	-5.222	0.000
Menza, 2010	% UAI events	0.828	0.771	0.889	-5.219	0.000
Mansergh, 2010	% UAI events	0.831	0.757	0.913	-3.884	0.000
Eaton, 2011	# sex partners	0.815	0.737	0.902	-3.960	0.000
Hightow-Weidman, 2012	# sex partners	0.818	0.742	0.901	-4.077	0.000
Hirshfield, 2012	% UAI events	0.833	0.755	0.919	-3.635	0.000
Koblin, 2012	% UAI events	0.843	0.761	0.933	-3.307	0.001
Mutanski, 2013	# UAI events	0.838	0.759	0.926	-3.460	0.001
Tobin, 2013	% UAI events	0.839	0.761	0.924	-3.558	0.000
Kurtz, 2013	# UAI events	0.840	0.767	0.920	-3.768	0.000
Harawa, 2012	# UAI events	0.846	0.776	0.922	-3.803	0.000
Lui, 2013	% UAI events	0.855	0.783	0.933	-3.508	0.000
Parsons, 2014	% UAI events	0.846	0.774	0.926	-3.637	0.000
Coffin, 2014	# UAI events	0.853	0.782	0.929	-3.621	0.000
Odonnell, 2014	# UAI events	0.859	0.790	0.933	-3.580	0.000
		0.859	0.790	0.933	-3.580	0.000

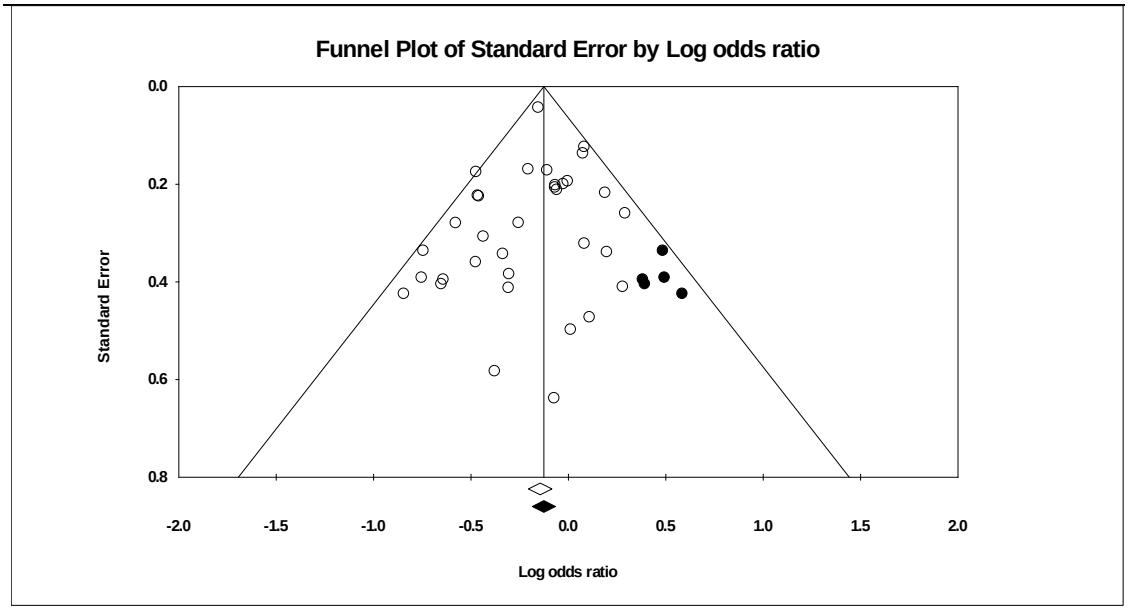


k=34, random effects model

Publication Bias Analysis

Analysis of publication bias indicated a small amount of risk based on observed asymmetry in the funnel plot (Figure 4). More studies are observed towards the top of the graph, and less towards the bottom. Slightly more trials were observed to the left of the mean effect size indicating there were a few trials that may be missing from the right side of the mean effect size (OR>1.0) due to reporting bias. Statistical analyses of potential publication bias suggests a small amount of bias.

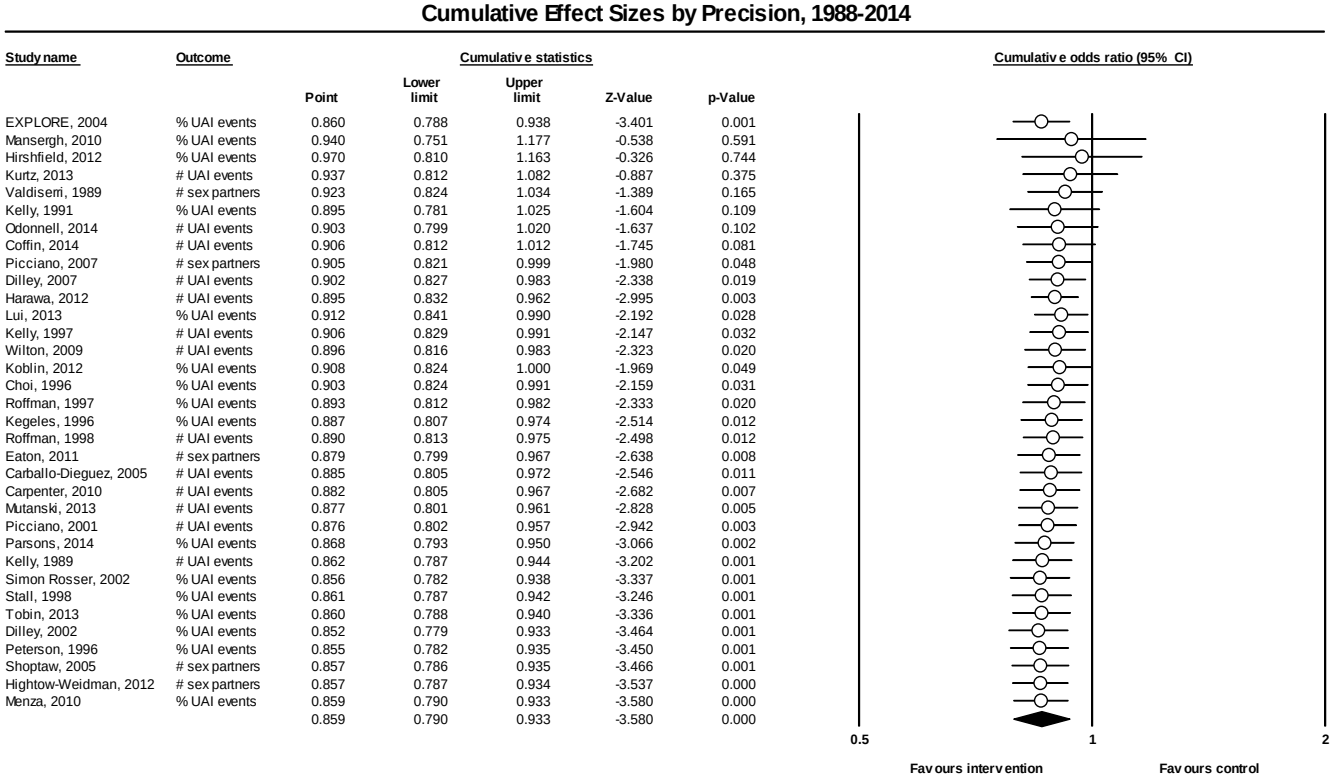
Figure 4
Funnel Plot for Publication Bias



Rosenthal's fail-safe *N* test required the addition of 143 missing trials to nullify the effect, or 4.2 studies per every observed trial. Orwin's *N* was more conservative, but indicated that 101 missing studies would be needed to bring the effect size over a mean odds ratio of 1.0. Both of these tests indicate minimal potential for publication bias. Trim and fill analysis suggested the imputation of five studies which aligns closely with our missing data procedures that resulted in 4 excluded studies. The filled circles in Figure 4 represent imputed values on the right side. Trim and fill

estimated the real effect size to be consistent with the observed effect size with *OR* = .858 (95% CI [0.790, 0.933]) under the random effects model, and *OR*=.865 under the fixed effects model (95% CI [0.814, 0.919]). Egger's regression was not significant (*B* = -0.35, *t*(32) = 1.19, 95% CI [-0.949, 0.247], *p* =0.120), indicating lack of publication bias. Finally, a cumulative meta-analysis of precision was conducted where trials were sorted by precision, and added in order of smallest to largest standard error. Visual examination did not detect a shift in overall effect size when the trials with larger standard errors (i.e. trials with smaller samples) were added, indicating minimal publication bias or minimal small-study effects. Results are presented in Figure 5. Synthesis of all publication bias analyses suggests this study was subject to minimal publication bias, though there may be some trials missing. Minimal bias results are consistent with literature cited previously that suggests statistically insignificant, and small studies, are likely to published in the HIV prevention literature.

Figure 5
Cumulative Effect Sizes by Precision



Outlier Examination

Only one trial produced a *z*-score higher than 3.0 and was subject to outlier examination. EXPLORE (Koblin et al., 2004) was the largest trial (*N*=4,295) with the smallest standard error (0.04) and *z*-score of -3.401 (OR=.860, 95% CI [0.788, 0.938]). Meta-analysis was repeated without EXPLORE and results were compared. Table 3 displays the mean effect sizes with all trials and then with EXPLORE removed. EXPLORE was retained in the overall analysis because its exclusion did not contribute to change in the overall result. EXPLORE was also retained due to its important contribution to HIV prevention as a landmark trial enrolling the largest group of MSM with the longest follow-up period. Sensitivity analysis using a "one study removed" analysis was performed to confirm that no single study (including EXPLORE) contributed to an observed shift in effect sizes. Forest plots produced from a one-study-removed analysis showed little shift in effect sizes (Figure 6).

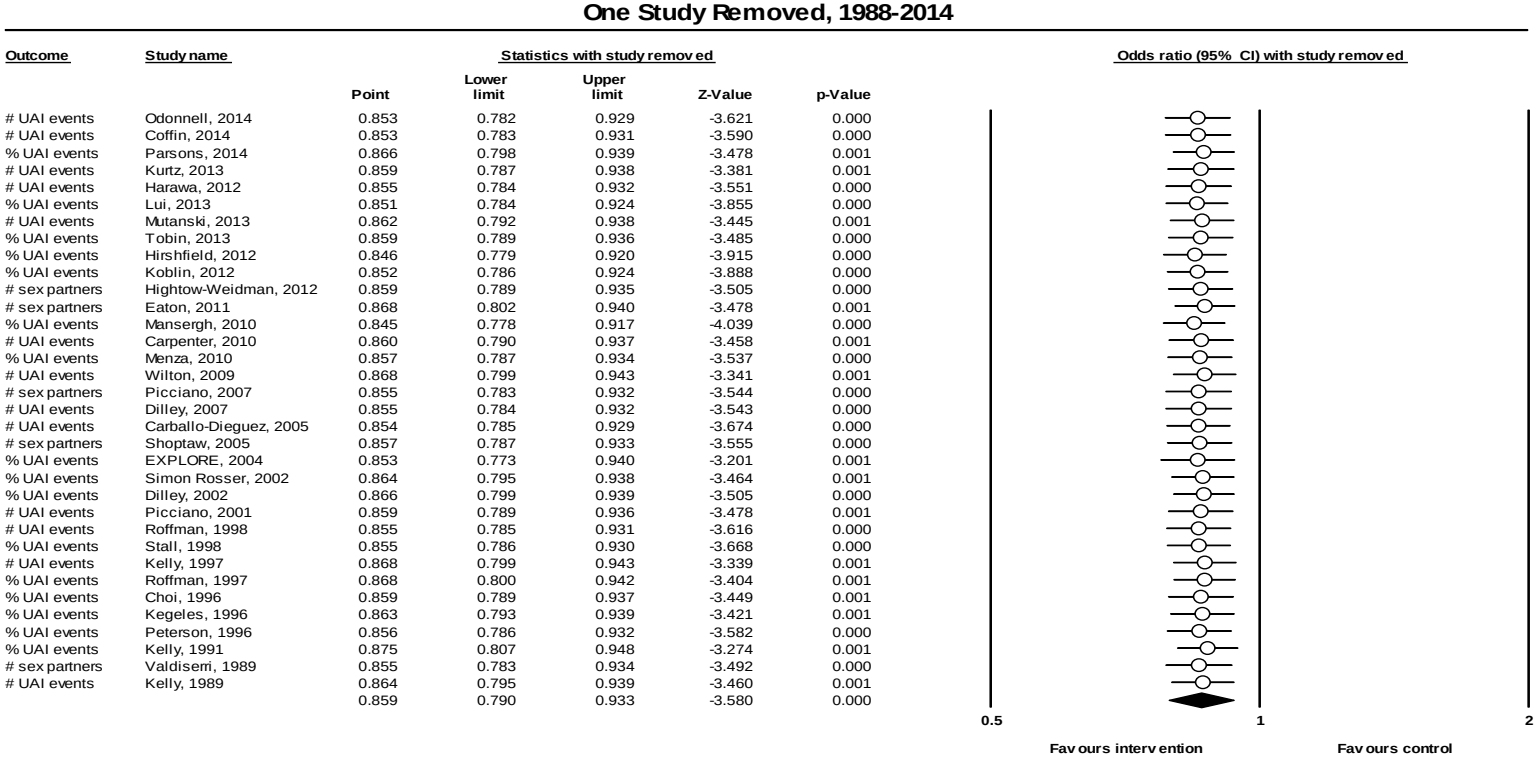
Table 3
Sensitivity Analysis

Analysis	*k*	Model	*OR*	95% CI		*P*	*Q (df)*
				LCL	UCL		
All studies	34	Fixed	0.865	0.815	0.919	<.001	39.35 (33)*
		Random	0.859	0.790	0.933	<.001	
EXPLORE removed	33	Fixed	0.870	0.800	0.946	<.001	39.32 (32)*
		Random	0.853	0.773	0.940	<.001	

Notes: *****not significant

Figure 6

One Study Removed Outlier Analysis



Moderator Analysis

The presence of moderating effects was not indicated by a significant *Q* statistic or an I^2 statistic that suggested heterogeneity. However, both *Q* and I^2 have limitations in interpreting heterogeneity, and due to the relatively small number of trials included, moderator analyses were conducted as planned. Effect sizes and moderator analyses identified *a priori* are displayed first in Table 4, followed by other moderator analyses using analog to ANOVA, and then meta-regressions. No variables hypothesized *a priori* as moderators were found to be significant (i.e. effect size measure, differential retention, intervention time span, and comparison group).

Two variables tested post-hoc were statistically significant at the $p<.05$ level in analog to ANOVA tests. Peer-led interventions were found to be more effective than non Peer-Led interventions (*OR*=0.676, 95% CI [0.569, 0.803], *Q*(1)=9.148, *p*=.002), Larger effect sizes were found for community-level interventions [*OR*=0.631, 95% CI [0.492, 0.808], *Q*(2)=6.879, *p*=.032), than individual or group-level interventions. Post-hoc meta-regressions of continuous variables revealed four statistically significant moderators ($p<.05$). The overall effect size decreased with year of publication (Q_R=5.80(1), *B*=.012, 95% CI [0.002, 0.022], *p*=.016), (Figure 7) supporting previous results from cumulative meta-analyses that showed declines in effect size over time. Effect sizes were larger when trials had a higher proportion of MSM under 30 or younger MSM (Q_R=16.02(1), *B*=.034, 95% CI [0.015, 0.055], $p<.001$), (Figure 8) and when trials had lower representation of MSM with a high school education or less (Q_R=4.49(1), *B*=.005, 95% CI [0.006, 0.009], *p*=.023) (Figure 9). Of trials that did not exclude HIV-positive persons, effect sizes were smaller

among trials with more HIV-positive participants (Q_R=5.50(1), *B*=.005, 95% CI [0.001, 0.008], *p*=.019) (Figure 10). HIV-status was further analyzed by creating a dichotomous variable to indicate when trials had more than 20 percent HIV-positive participants. Based on this analysis, trials with less HIV-positive participants were observed to have larger effects (*OR*=.809, 95% CI [0.726, 0.902], *Q*(1)=4.58, *p*=.032). Finally, 11 trials excluded HIV-positive participants in their eligibility criteria. In subgroup analyses, trials that excluded HIV-positive MSM had larger effects (*OR*=.765, 95% CI [0.640, 0.916], *Q*(2)=9.742, *p*=.008).

ANOVA analog also indicated trend level ($p < .10$) differences between Evidence-Level [*Q*=6.869(3), *p*=.076]. Trials classified as highest-level evidence, Evidence-Based Interventions (EBI), showed larger effect sizes (*OR*=.776, 95% CI [0.682, 0.883]) than lower levels of evidence. Trials focusing on specific MSM subpopulations suggested varying effects based on subpopulation [*Q*=6.271(3), *p*=.099). Trials focused on young MSM (*OR*=.616, 95% CI [0.418, 0.907]) and trials with no specific population (*OR*=.825, 95% CI [0.793, 0.913]) appeared to be more effective than trials focusing on substance users or MSM of color. Finally, trials with <80 percent overall retention appeared to be slightly more effective than trials with higher retention (*OR*=.775, 95% CI [0.667, 0.901], *Q*(1)=2.809, *p*=.094).

Table 4
Moderator Analyses

Variable	*K*	*OR*	*SE*	95% CI		*P*
				LCL	UCL	
Mean Effect Size						
Fixed	34	0.865	0.026	0.815	0.919	.<001
Random	34	0.859	0.036	0.790	0.933	<.001
Homogeneity	*Q* [33] = 39.350, *p* =.207					
Comparison Condition*						
HIV-related	17	0.857	0.012	0.787	0.932	.666
Non HIV-related	7	0.947	0.068	0.728	1.232	
Wait-list	10	0.813	0.052	0.661	0.998	
Effect Measure*						
Count-level	16	0.859	0.028	0.742	0.993	.835
Dichotomous	18	0.842	0.051	0.167	0.369	
Retention Better*						
Control	10	0.857	0.017	0.794	0.926	.861
Equal	24	0.846	0.114	0.745	0.961	
Time Span*						
≤1 month	20	0.847	0.022	0.754	0.952	.792
>1 month	14	0.868	0.022	0.758	0.993	

Table 4
Moderator Analyses

Variable	*K*	*OR*	*SE*	95% CI		*P*
				LCL	UCL	
Pre or Post ART						
Pre	11	0.866	0.004	0.812	0.968	.273
Post	23	0.792	0.022	0.661	0.949	
MSM Subpopulation						
Young MSM	4	0.616	0.133	0.418	0.907	.099
MSM of color	9	0.881	0.718	0.742	1.056	
Substance Users	6	0.996	0.842	0.842	1.179	
None	15	0.825	0.729	0.793	0.913	
Region						
Northeast	7	0.887	0.056	0.703	1.118	.523
West	13	0.859	0.039	0.729	1.013	
South	4	0.749	0.041	0.605	0.927	
Midwest	2	0.577	0.208	0.340	0.979	
National	4	0.809	0.118	0.560	1.170	
Multi-site	4	0.924	0.037	0.751	1.137	
Intervention Level						
Individual	14	0.872	0.019	0.795	0.981	.032
Group	17	0.909	0.023	0.806	1.025	

Table 4
Moderator Analyses

Variable	*K*	*OR*	*SE*	95% CI		*P*
				LCL	UCL	
Community	3	0.631	0.053	0.492	0.808	
Multiple Outcomes						
Yes	9	0.915	0.030	0.785	1.066	.372
No	25	0.842	0.017	0.760	0.933	
Significant Result						
Yes	21	0.826	0.016	0.744	0.916	.217
No	13	0.921	0.026	0.801	1.059	
≥80% Retention						
Yes	19	0.898	0.013	0.826	0.975	.094
No	15	0.775	0.032	0.667	0.901	
Follow-Up Assessment						
≤3 Months	11	0.841	0.038	0.709	0.998	.796
>3 Months	23	0.864	0.015	0.782	0.955	
Peer Delivery						
Yes	7	0.676	0.033	0.569	0.803	.002
No	27	0.901	0.012	0.840	0.968	
Sexual Communication						
Yes	23	0.838	0.013	0.764	0.919	.228

Table 4
Moderator Analyses

Variable	*K*	*OR*	*SE*	95% CI		*P*
				LCL	UCL	
No	11	0.954	0.047	0.789	1.154	
Skills Building						
Yes	18	0.876	0.011	0.819	0.937	.433
No	16	0.813	0.043	0.684	0.967	
Individual Plans						.118
Yes	27	0.829	0.017	0.749	0.917	
No	7	0.965	0.030	0.821	1.133	
Stigma						.997
Yes	13	0.854	0.019	0.791	0.922	
No	21	0.854	0.025	0.754	0.967	
Technology						.704
Yes	8	0.826	0.048	0.673	1.014	
No	26	0.863	0.015	0.785	0.949	
Evidence Level						.076
EBI	11	0.776	0.020	0.682	0.883	
Positive	10	0.839	0.041	0.846	1.368	
Rigorous	7	0.952	0.040	0.700	1.007	
Other	6	1.076	0.064	0.784	1.157	

Variable	*K*	Q_R	*B*	*SE*	*B* 95% CI		*P*
					LCL	UCL	
Study Year	34	5.08	0.012	0.005	0.002	0.022	.016
Demographics							
Age	34	16.02	0.034	0.010	0.015	0.055	<.001
% Black	29	0.03	-0.000	0.001	-0.003	0.003	.864
% White	29	0.03	-0.000	0.001	-0.003	0.002	.854
% Latino	29	0.75	0.002	0.002	-0.002	0.005	.387
% Gay or Bisexual	20	0.93	0.008	0.009	-0.009	0.025	.335
% HIV-positive	23	5.50	0.005	0.002	0.001	0.008	.019
% High school or less	25	4.49	0.005	0.002	0.006	0.009	.034
Number of Sessions	20	0.01	0.001	0.010	-0.018	0.021	.910
Number of Hours	18	0.01	0.001	0.007	-0.013	0.014	.934

*=moderators hypothesized *a priori*

Figure 7
Scatter Plot of Log Odds Ratio by Year Published

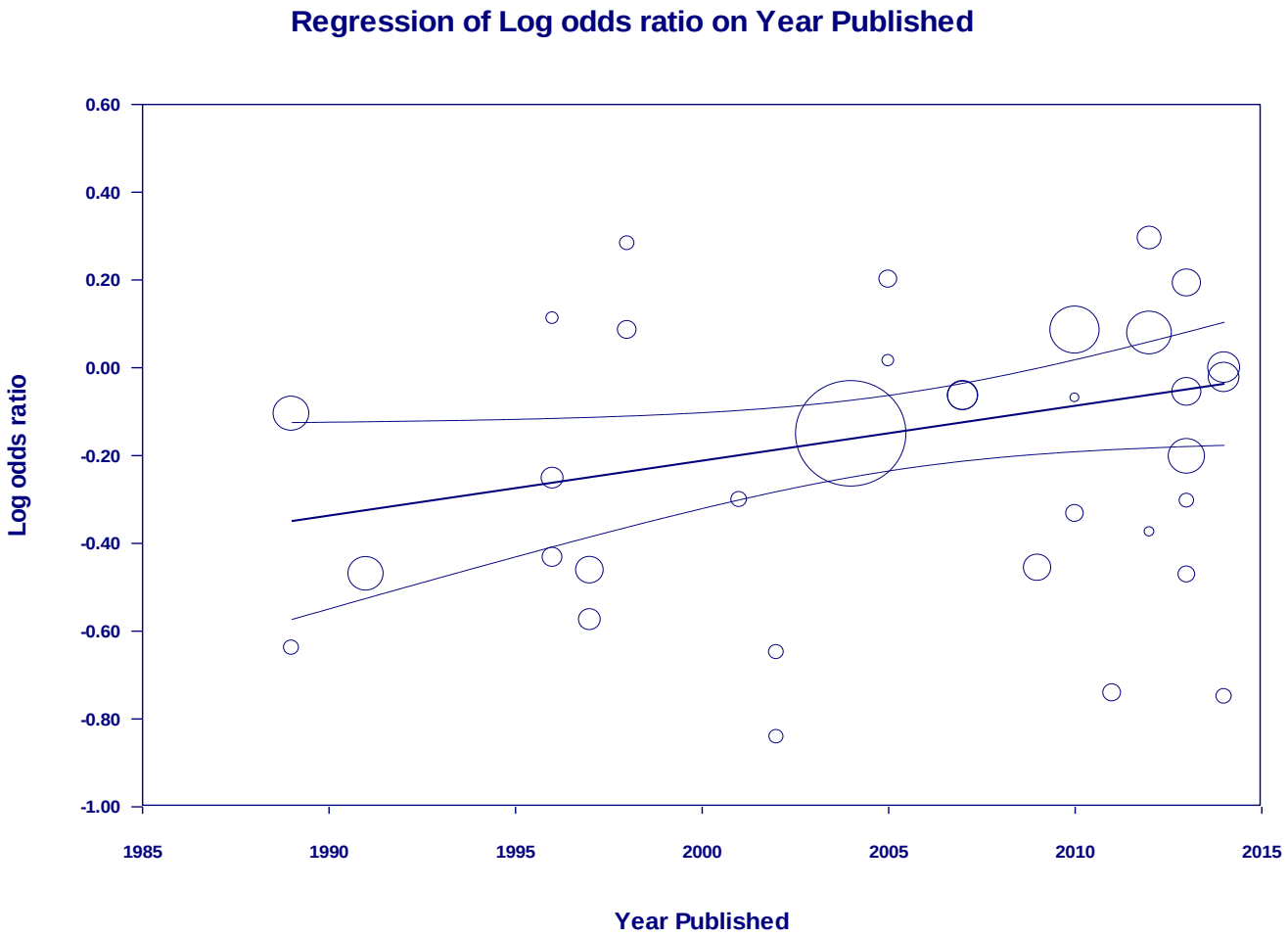


Figure 8
Scatter Plot of Log Odds Ratio by Age

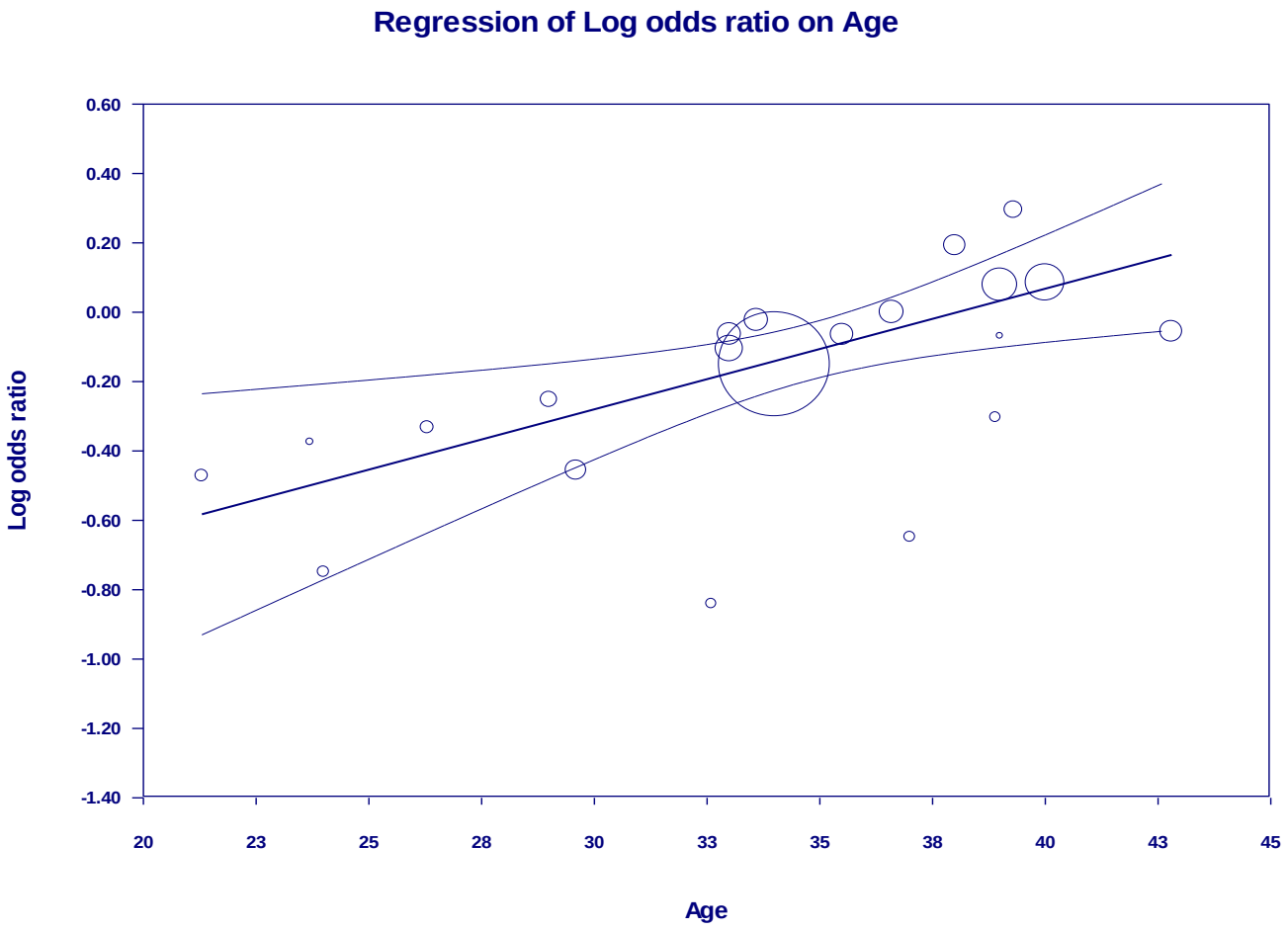


Figure 9

Scatter Plot of Log Odds Ratio by Education Level

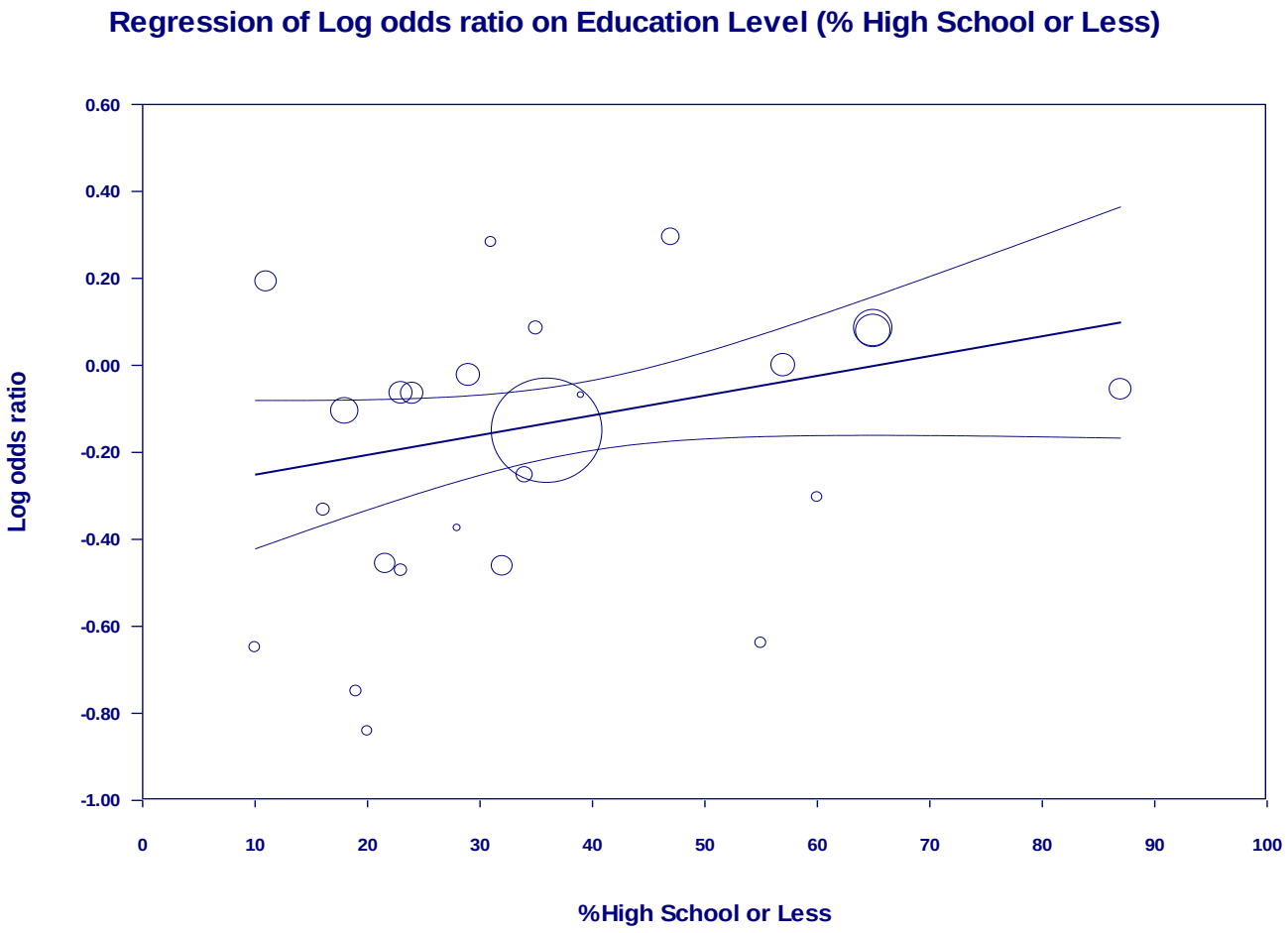
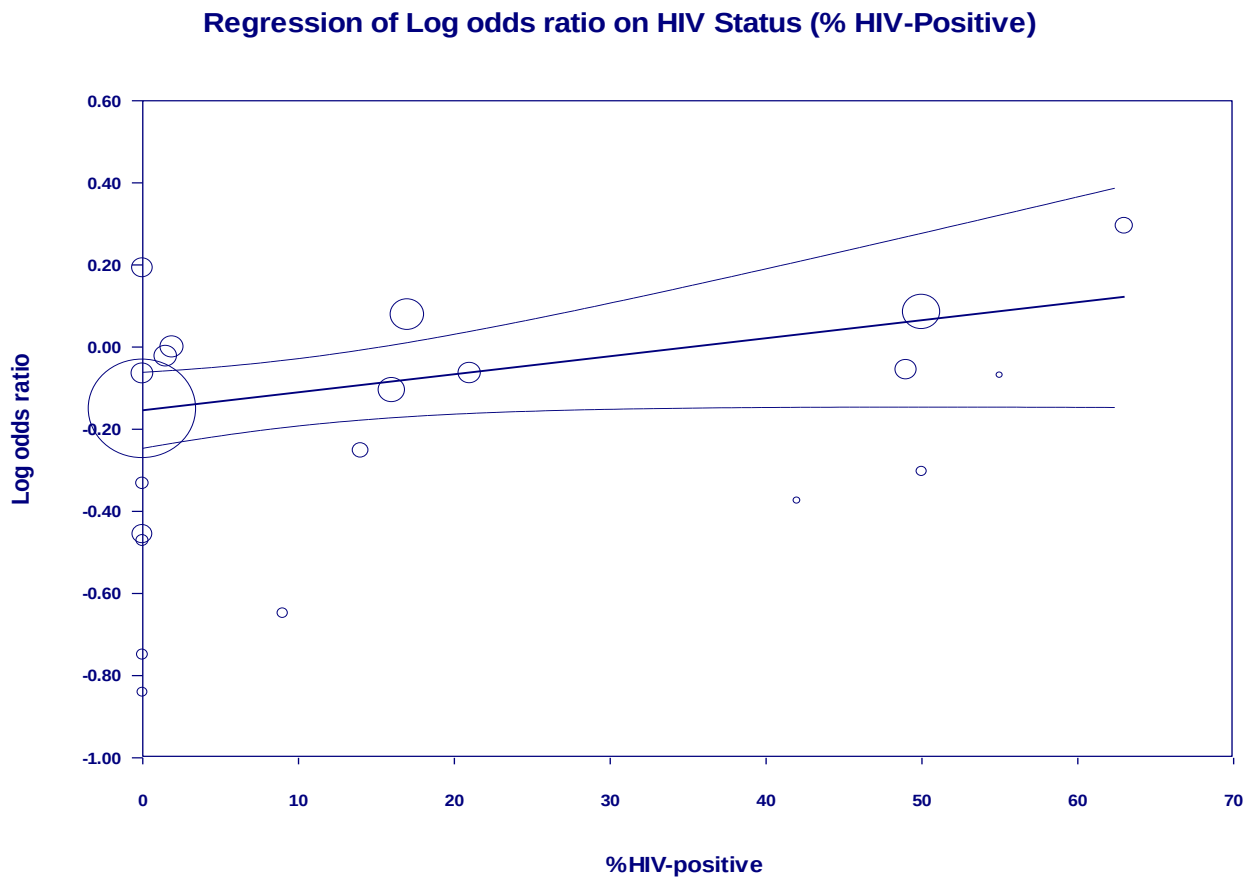


Figure 10
Scatter Plot of Log Odds Ratio by HIV Status



CHAPTER 5: DISCUSSION

Summary of Findings

This study evaluated effects of 34 randomized controlled trials of HIV behavioral interventions for 17,872 MSM conducted between 1989 and 2014. Overall, behavioral interventions reduced the odds of sexual risk behavior by 14 percent (*OR*=.859, 95% CI [0.790, 0.933], $p<.001$). These findings show that behavioral interventions are still somewhat effective to prevent HIV transmission among MSM. However, the updated effect size was considerably smaller than the magnitude of effects observed in earlier meta-analyses.

This study also examined effect size moderators, integrated interventions, and cumulative effect sizes over time. Intervention effects were highly statistically homogenous, and all moderators hypothesized *a priori* were insignificant. Post-hoc analyses found eight variables (subpopulation, retention, peer delivery, evidence-level, intervention-level, age, HIV status, education) to be statistically related to effectiveness, but findings need to be interpreted with caution due to their exploratory nature. Qualitative review identified nine trials as integrated interventions; six trials included a primary outcome related to substance use and three included outcomes related to HIV testing. Consistent with study hypotheses, most of these trials were conducted after the last review in 2007.

Most notably, this study revealed an unexpected finding related to the cumulative effect size over time. Cumulative meta-analysis revealed that intervention effects gradually weakened over time. From 1991 to 2014, the magnitude of the effect

size decreased by 19.5 percent (*OR*=.719-.859). Follow-up meta-regression analysis using publication year as a moderator variable added statistical evidence for effect size shrinkage over time. Reasons for effect size decline were not able to be explained due to statistical homogeneity. However, several factors such as HIV prevention fatigue, inclusion criteria, comparison condition, and underpowered trials may contribute to effect size shrinkage.

Overall Meta-Analysis and Effect Size

This meta-analysis provides updated evidence that HIV behavioral interventions still show significant overall effects to reduce sexual risk behavior among US MSM. However, the current evidence is weaker than hypothesized. Findings support current thinking that behavioral interventions are necessary, but no longer sufficient for HIV prevention (Coates, 2013). To our knowledge, this meta-analysis is the most recent quantitative synthesis of behavioral interventions designed for US MSM, and includes trials conducted up to the end of calendar year 2014. Multiple tests were conducted to support the validity of its overall finding. Publication bias analyses indicated little to no publication bias, and small study effects were not shown to inject bias when effect sizes were plotted by precision. Iterative sensitivity analyses did not reveal any significant outliers. These results combined with previous meta-analyses provide good support for the continued funding of behavioral intervention trials, especially when implemented as part of combination prevention agendas.

The observed 14 percent decrease in sexual risk behavior is modest, and considerably smaller than the last meta-analysis. While a 14 percent decrease may still confer some protection on an individual level, significant changes in HIV epidemics require larger effect sizes maintained over time, and among more people (Coates et al., 2008). More successful interventions have decreased the odds of risk behavior by about 25 percent (Johnson et al., 2002). Even if HIV behavioral interventions were to demonstrate greater reductions in risk behavior, it is unlikely that they can avert large numbers of infections (IOM, 2001).

This study hypothesized that the updated effect size would be comparable to previous meta-analyses, despite variations in study protocols. Previous meta-analyses found behavioral interventions to decrease sexual risk behavior by about twice as much (Herbst, et al., 2005; Johnson et al., 2002a; Johnson et al 2005; Johnson et al., 2008). In the last meta-analysis, Johnson (2008) found interventions to decrease UAI by 27 percent when compared to non-HIV related controls, and by 17 percent when compared with HIV-related controls. While it is possible that this study's scope was too different to merit meaningful comparisons to other meta-analyses, it is clear that effects have changed over time. Observed effect size shrinkage warrants additional research including replication studies to examine if earlier trials (especially pre-ART) are no longer comparably effective among current MSM populations (Higa et al., 2013).

Gradual Effect Size Shrinkage

Contrary to the hypothesis, this study found that the effect size was only about half as large as expected. Further, cumulative meta-analysis showed that the effect size gradually weakened over time, and has slowly moved closer to the null value of 1. Visual results were supported by meta-regression results that revealed a linear relationship between publication year and effect size; the overall effect size declined with each new publication year (Figure 7). Reasons for effect size decline are not known, but may be related to multiple factors including MSM subpopulation, HIV prevention fatigue, comparison condition, and underpowered trials.

Cumulative meta-analysis (CMA) was used to accomplish the study objective of displaying effect sizes over time. CMA addresses the impact of each new trial on prior pooled results and assesses how robust meta-analytic results remain over time (Trikalinos et al., 2004; Ioannidis, Contopoulous-Ioannidis & Lau, 1999). A systematic review of over 1,500 CMAs conducted on healthcare interventions discovered three main patterns of effect size change: early positive results became null or negative over time; null or negative results became positive over time; or results stabilized, but intervention research continued (Clarke, Brice, & Chalmers, 2014).

Studies demonstrating effect size shrinkage over time with CMA are not uncommon, and CMA may be particularly important in fields where small trials dominate the RCT literature (Bollen, Utterwaal, & Vaught, 2003; Hanson & Broom, 2005; Klein, Jacobs, & Reinecke, 2007; Zhang et al., 2010; Trikalinos et al., 2004). Trikalonas et al. conducted a review of 100 meta-analyses of mental health

randomized trials (*N*=99,303) to test how effect sizes changed over time. The authors calculated the size and direction of the relative change in the cumulative effect size at each calendar year. Relative changes were calculated by the formula: $OR_{subsequent\ step}/OR_{current\ step}$. Relative changes over 1 indicated shift favoring the comparison over intervention group. They found that the magnitude of the effect size shifted considerably from early positive results to null results, raising questions about how effect size shrinkage in the mental health field may impact declining clinical importance.

In this study, the most noticeable shift in effect size occurred with the inclusion of EXPLORE in 2004 when the cumulative ES shifts from .720 to .777. Using Trikalinos's calculations, this shift equates to a relative change over 1 and a shift unfavorable to the intervention group (Trikalonas et al., 2004). After 2005, the effect size clearly stabilizes, showing gradual effect size shrinkage over time. The impact of EXPLORE's weight is the most likely explanation for the overall effect size shift; EXPLORE is the largest behavioral trial to date (*N*=4,295, *OR*=0.86) and accounts for the largest relative weight in this study (18.2). Generally, HIV prevention research is dominated by smaller studies. Median sample size for this study was 328 and only two trials in this study had samples over 1,000 participants. In addition to its unusually large sample size, EXPLORE was also the only trial with a *z*-score >3.0. Since sensitivity analyses did not show that EXPLORE exerted influence on the overall effect size, it was included in the meta-analysis. While the weight of EXPLORE may have contributed to some effect size shift, there are likely other

important reasons that contribute more to effect size decline. In many cases of CMA, there are limited data to fully understand effect size shift over time (Iaonnidis et al., 1999).

To our knowledge, CMA has not been conducted to assess the robustness of effect sizes for HIV behavioral interventions over time. Effect size shrinkage may reflect contextual or temporal factors that influence intervention effectiveness (e.g., HIV prevention fatigue or reduced fear of HIV post-ART), trends in experimental design (e.g. MSM subpopulation, choice of comparison group), or underpowered trials. Potential limitations of current behavioral interventions have been cited in other studies and argue that behavioral interventions, when implemented alone, are outdated (Higa et al., 2013; Wolfeiler & Ellen, 2007; Kippax, 2007; Ross & Wight, 2007; Coates et al., 2008; Coates, 2013; Sullivan et al., 2012). This study aimed to identify the most effective components of behavioral interventions to inform next-generation combination approaches. Due to statistical limitations related to homogeneity, this study used qualitative review to examine potential limitations of interventions.

Effect of Underpowered Trials

A priori power analysis was not performed for this meta-analysis, and investigation of post-hoc power was not a study objective. Additionally, assessing post-hoc power for the primary studies was out of scope for this review, but is a relevant question for future reviews. Meta-analyses will often have higher power than the primary studies due how to precision changes in a meta-analysis (Borenstein et al.,

2009). Trialists across various fields are encouraged to conduct *a priori* power analyses to ensure at least 80 percent power to detect an effect. However, many trials do not report *a priori* power analysis for sample specification, or do not achieve 80 percent power upon post-hoc analysis (Borenstein et al., 2009). Underpowered trials are a particular problem for intervention research using small samples. Reviews of the psychological literature have demonstrated improved overall power in trials over time, but trials examining small effects remain persistently underpowered (Rossi, 1990; Maddock & Rossi, 2001).

A recent review of 14,866 meta-analyses in 1,991 Cochrane reviews calculated power per study in each meta-analysis and found that 70 percent of Cochrane meta-analyses were underpowered (Turner, Bird, & Higgins, 2013). Of the 34 trials reviewed in this study, less than one third of trial manuscripts (k=10) reported an *a priori* power analysis at 80 percent for primary outcomes. Lack of reporting of power analysis does not necessarily imply that one was not performed or that power was low, but highlights the need for more transparent reporting in RCTs. Many studies did not specify the primary outcome, and results extracted for meta-analysis may have been for a secondary outcome that was not adequately powered, and thus not appropriate for meta-analysis (Turner et al., 2013).

Finally, CDC efficacy criteria specify a minimum threshold of 40 subjects per intervention and comparison arm (CDC, 2014c) to be classified as an evidence-based intervention. Most trials in this study met this criterion, and only four did not. Further review of RCTs for evidence-based behavioral interventions may benefit from

post-hoc analyses similar to Turner et al., 2013 and Rossi (1990) to determine if a general guideline of 40 participants per arm adequately assesses power for the primary study outcome.

Post-ART and HIV Prevention Fatigue

This study found that most trials (*k*=23) were conducted after the introduction of antiretroviral therapy (HIV medications) in 1996 (post-ART). Eleven trials were conducted pre-ART. Trials conducted later in time are more likely to face contextual challenges related to HIV prevention fatigue or ART-optimism. ART-optimism is when high-risk individuals become less concerned about HIV infection because of the disruptive success of ART; HIV prevention fatigue describes the attitude that HIV prevention has become tiresome and leads to fatigue in maintaining safer sex behaviors over time (Ostrow et al., 2002; Stockman et al., 2004).

Distribution of pre- and post-ART trials in this study are different than in Johnson's 2008 review. This is mostly due to more MSM-specific trials conducted after 1996 (82 percent), and a high number conducted 2010 or later (44 percent). Johnson's review protocol resulted in trials being about equally distributed between pre- and post-ART years (24 and 23 respectively). While neither meta-analysis found pre- or post-ART to be a significant moderator of effect, early pre-ART interventions were associated with the most successful examples of behavior change resulting in dramatic decreases in HIV incidence (Coates et al., 2008; Wohlfeiler & Ellen, 2007). Within 10 years, this effect has decreased (Wohlfeiler & Ellen 2007). It is not known if pre-ART trials, or even trials older than 5-10 years, would be effective among

current MSM. This observation raises important questions about the need for replication studies of all trials deemed to be evidence-based, especially when these trials used small samples, but continue to be recommended in public health practice.

Inclusion Criteria or MSM Subpopulation

Higa (2013) suggested that HIV prevention research is more now challenging than previously due to an increased focus on enrolling the highest risk MSM for behavioral interventions. Current recommendations suggest targeting MSM at the highest risk of HIV acquisition and transmission. As many other MSM of lower risk have benefited from behavioral interventions, the remaining at-risk MSM subgroups continue to challenge the limits of current interventions due to complex risk factors not easily addressed by short-term interventions. Priority MSM subgroups include substance using MSM, black and Latino MSM, young MSM, HIV-positive MSM, and sexually high-risk MSM. All priority MSM subgroups experience "syndemics." Syndemics are defined by "a set of enmeshed and mutually enhancing health problems that *working together* in a context of deleterious social and physical conditions increase vulnerability, significantly affect the overall disease status of a population" (Singer, 2010). Current behavioral interventions generally cannot address structural or other risk factors that contribute to syndemics such as poverty, stigma and discrimination, health care access, and mental health. These unaddressed risk factors contribute to ongoing sexual risk, and interventions demonstrated to be effective in the research setting are unlikely to produce sustained effects due to social or economic inequities that persist beyond the intervention period.

Comparison Condition

Johnson (2008) found interventions to decrease unprotected anal intercourse by 27 percent when compared to non-HIV related controls, and by 17 percent when compared with HIV-related controls. This study found that effect size did not statistically differ by control condition. Interventions using non-HIV related controls decreased UAI by six percent, interventions using HIV-related controls decreased UAI by 14 percent, and interventions using wait list controls decreased UAI by 19 percent. Most trials used an HIV-related control (k=17) or a wait-list control (k=10), and only seven trials used non-HIV related control. In Johnson's review, 65 percent of trials used non-HIV related controls. This discrepancy suggests that more recent trials designed exclusively for MSM are more likely to choose an HIV-related control. This observation is supported by Higa et al., (2013) who found that wait-list controls decreased over time, and demand controls using an HIV-related comparison increased.

While there are certain advantages to using a wait-list or non-HIV related controls for rigorous evaluation (Menza et al., 2010), others argue that use of wait-list or non-HIV related attention controls is unethical because it withholds potentially effective treatment conditions from high-risk individuals (Higa et al., 2013). However, use of better control conditions that essentially use a diluted version of the intervention may greatly reduce the trial's ability to detect effects (Crepaz et al., 2015; Higa et al., 2013). This study found that 21 trials did not find significant results between groups; however many of these trials did report statistically significant risk reduction changes in both intervention and comparison groups. While this may be

attributed to assessment bias, it is also possible that the value of exposure to minimal or standard HIV-related intervention may be underestimated due to the lack of finding a "significant positive effect" between groups. HIV prevention research may benefit from a standardized comparison condition to compare intervention effects across trials (Crepaz et al., 2015; Higa et al., 2013), as well as further examination into control group effects.

Integrated Interventions

Only nine trials were classified as "integrated interventions." Integrated interventions were defined as interventions that simultaneously address multiple problem behaviors that share a root cause (e.g. substance use, mental health, homelessness, stigma) (Crepaz et al., 2015). This study found that integrated interventions were more common among trials designed for substance users (k=6); these trials included at least one outcome (other than sexual risk reduction) directly related to reducing substance use (Harawa et al., 2012; Menza et al., 2010; Parsons et al., 2014; Picciano et al., 2001; Picciano et al., 2007; Shoptaw et al., 2005). The other three integrated interventions included additional outcomes related to the receipt of HIV tests at follow-up (Hirshfield et al., 2012; Odonnell et al., 2014; Wilton et al., 2009). Of trials with HIV testing outcomes, two were designed for black or Latino MSM. Consistent with the study hypothesis, most integrated interventions (k=6) were conducted after the last review in 2007 and suggests an increased use of these approaches for MSM.

Integrated interventions are an emerging research priority in the development of combination prevention approaches for high-risk groups. In 2015, CDC published a systematic review and meta-analysis that showed positive effects of integrated interventions for people living with HIV (Crepaz et al., 2015). Their review found that integrated interventions were effective in reducing sex without condoms, and showed promise for increasing adherence to HIV medications. At the 2015 National HIV Prevention Conference, CDC announced that at least two EBIs for MSM had been adapted to an integrated intervention approach (Collins, 2015). For example, *Many Men, Many Voices (3MV)* integrated a new component specific to pre-exposure prophylaxis (PrEP) for black MSM, and *Popular Opinion Leader* has integrated multiple outcomes related to HIV care as well as PrEP initiation and adherence. However, trial results from these adaptations have not been shared or published.

To our knowledge, there is no systematic review or research synthesis describing the effects of integrated interventions among high-risk HIV-negative MSM. This study aimed to understand research trends in using this approach, and hypothesized that more integrated interventions would be available for review than in the past review (pre-2008). This study also examined integrated interventions as a moderator variable and found that integrated interventions were not a significant moderator of effect ($p>.05$). However, five of nine trials reported statistically significant effects on at least one sexual risk behavior outcome (Harawa et al., 2012; Hirshfield et al., 2012; Odonnell et al., 2014; Parsons et al., 2014; Wilton et al., 2009), and four of the nine integrated interventions were also classified as EBIs (Harawa et

al., 2012; Odonnell et al, 2014; Parsons et al., 2014; Wilton et al., 2009). Demonstration of positive effects among hard-to-reach MSM subpopulations such as substance users and MSM shows promise for using integrated interventions in combination prevention approaches, but further research beyond this review is needed.

While this study can offer some descriptive information from its narrative review, further research synthesis including data from recent EBI adaptations are required to determine if integrated interventions are effective in reducing sexual risk behavior among high-risk MSM. As stated previously, most integrated interventions were conducted after 2008 and this area of research is relatively young. Integrated interventions have intuitive appeal due to their logical approach in addressing syndemic factors of HIV risk (Parsons, Grov, & Golub, 2012; Starks, Miller, Eggleston, & Parsons, 2014). Integrated interventions may also have implementation benefits since they would likely require fewer intervention sessions than single-target interventions that address one behavior at a time (Crepaz et al., 2015). However, it is not yet known if integrated interventions are more effective than single-target interventions, and if there would be negative effects related to scaling up integrated interventions. Single-target interventions provide clear evidence about what works when changing one behavior at a time, and it is not clear if integrated interventions may dilute any single outcome (Crepaz et al., 2015). Updated reviews that examine the most recent examples of integrated interventions applied in the field are necessary to answer these questions.

Moderator Analyses

Overall, trials seemed to have considerable clinical heterogeneity due to their diverse intervention components, control conditions, and design variables. However, an insignificant *Q* statistic combined with an I^2 value less than 25 percent suggested very little statistical heterogeneity. Heterogeneity this low was not expected, but homogeneity was also observed in the previous meta-analysis (Johnson et al., 2008). Lack of observed heterogeneity was disappointing, and restricted this study's overall goal to identify factors statistically associated with intervention effectiveness. Lack of heterogeneity generally indicates that moderator analyses are not warranted, unless they are hypothesized *a priori*. However, typical measures of heterogeneity have some level of uncertainty, or in other words, "lack of evidence of heterogeneity is not evidence for homogeneity" (Borenstein, Higgins, Rothstein, & Hedges, 2015; Ioannidis, Patsopoulous, & Evangelou, 2007). Following Johnson et al (2008)'s example, this study pursued post-hoc analyses despite statistically homogeneous effects. While results need to be interpreted with caution due to potentially spurious findings, these data can inform hypotheses for future reviews.

This study used subgroup analyses and meta-regressions to explore independent variables previously examined by CDC's Prevention Research Synthesis reviews. Twenty-two analyses were conducted in total, and four independent variables were specified *a priori. A priori* moderators included comparison condition, retention, time span, and outcome measure. All hypothesized moderators were insignificant at the $p<.05$ level (p-values ranged from $p=.67$ to $p=.86$). This finding

was surprising. Eighteen other variables were examined post-hoc. Five were significant at the $p<.05$ level (e.g. age, HIV status, education, peer-led, and intervention-level) and three were significant at the trend level of $p<.10$ level (e.g. evidence-level, subpopulation, and retention). All independent variables found to be significant are supported by context and previous research, and thus warrant further examination in future meta-analyses. Moderators are presented in order of relevance to the study's scope.

Evidence Level

Trials classified as evidence-based interventions (EBIs) as opposed to Rigorous Non-EBIs, Positive Non-EBIs, and Other Non-EBIs were shown to be most effective at the trend level (*OR*=.776, *p*=.076). Currently, there are 20 EBIs listed for MSM out of 93 other EBIs in CDC's Compendium for Evidence-Based Interventions (CDC, 2014b). Eleven are for HIV-negative MSM, and nine are for HIV-positive MSM. All 11 EBIs for HIV-negative MSM were included in this review and coded as EBIs. Ten trials were classified as Positive-Non EBIs (i.e. significant finding, but did not meet criteria for methodological quality), seven trials were classified as Rigorous-Non EBIs (i.e. met methodological criteria, but did not find a significant result), and six trials were classified as Other Non-EBIs.

Consistent with findings from Higa et al. (2013), all Rigorous Non-EBIs were conducted post-ART, suggesting improved methodological quality over time. When Rigorous Non-EBIs were compared to EBIs, they were more likely to include higher-risk MSM with multiple vulnerabilities (e.g., substance users and low education).

Trials enrolling populations with higher baseline risk, and more barriers to behavior change have greater challenges to demonstrating efficacy compared to trials enrolling more compliant MSM of low to moderate risk.

Of the 10 trials classified as Positive Non-EBIs, the most frequent reasons for not achieving EBI or Rigorous classification were: having analytic samples under 40 participants per condition (*k*=2), having only one immediate follow-up session in both comparison groups *(k*=3), having less than 60 percent retention in either comparison (*k*=2), or having other methodological fatal flaws (*k*=3) such as contradictory findings, considerable missing data, or reassignment of participants. Interestingly, five out of seven interventions using technology (i.e. internet, telephone) were classified as Positive Non-EBIs. Technology-based interventions struggle with high loss to follow up rates, and most did not meet EBI criteria for retention. Other Non-EBIs did not find a positive effect and they did not meet criteria for methodological quality. Similar to positive non-EBIs, methodological limitations included analytic samples with less than 40 participants per condition (*k*=2), only one immediate follow-up session *(k*=2), or having other fatal flaws (*k*=2).

Age

Trials with higher proportions of younger participants were associated with greater intervention effects ($p<.001$). Mean age of all trials was 33.3 (*SD*=5.4). Scatter plot of meta-regression result showed that intervention effects declined with increasing mean age (Figure 8). The effect of age was supported by a separate subpopulation moderator analysis that examined trials designed for high-risk groups,

including young MSM. Young MSM were defined as MSM under 30 years of age. Trend-level results were found for the overall group comparison (p=.09), and post-hoc analyses revealed that interventions for young MSM (OR=.616) were associated with larger effects than all other subgroups.

While these findings are hopeful for future trials targeting young MSM, these findings highlight an important HIV prevention research gap. Only four trials were designed for young MSM. In the US, HIV infections have increased most among young MSM aged 13-24. Further, young black MSM account for more new HIV infections than any other age or race group, and it is estimated that at least 50 percent of young black MSM will be HIV-positive by the time they turn 35 (CDC, 2014a; Stall et al., 2009). These results show that younger MSM are willing to participate in behavioral interventions, and can be effectively recruited for HIV prevention research. Younger MSM who have had less exposure to behavioral interventions may be more amenable to sexual risk behavior change than older MSM who experience "HIV prevention fatigue" (Sullivan et al., 2012). However, there may be other challenges associated with enrolling young MSM. Trials for younger MSM had smaller sample sizes than the median sample size (range: 39-113) and retention was lower among this group compared to other MSM. Only one of four trials for young MSM achieved an overall retention rate ≥80 percent; the three other trials had retention rates between 56 percent and 79 percent.

Subpopulation

A four-level independent variable was created to examine the amount of research allocated to high-risk MSM subpopulations, and then to examine if effects varied by subpopulation. Subpopulations included young MSM, substance users, MSM of color, and none. As stated previously, overall results indicated trend-level significance (p=.09). Interventions designed for substance users (k=6) or MSM of color *(*k=9) were less effective than interventions for young MSM (k=4) or none (k=15).

Over the past ten years, there has been an increase in interventions designed specifically for black and Latino MSM in an effort to align research with the current epidemic. Despite an increased research focus, there are few evidence-based interventions for MSM of color. To date, there are only four EBIs for MSM of color in CDC's Compendium; two are for black MSM, one is for Latino MSM, and one is for Asian/Pacific Islander MSM. As the epidemic shifts to become increasingly concentrated among Black and Latino MSM, well-funded research agendas will be essential to develop and test behavioral interventions that work for MSM of color. Future research in this area should aim to develop interventions guided by a comprehensive approach to address the social, cultural, and structural factors (i.e. economic or social inequities) that may influence effectiveness (Sullivan et al., 2012), as well as factors related to research participation (Hatfield et al., 2010).

Trials focusing on substance-using MSM (SUMSM) have also increased in the past decade, yet there is only one EBI for SUMSM and it was added in 2015 (Parsons

et al., 2014). Substance use is well-known to be an important driver of HIV infection among MSM (Coates et al., 2008; Sullivan et al., 2012). This study found that interventions were less effective among substance-users than any other groups. Failure to find significant effects highlights the persistent HIV prevention research challenge of recruiting, retaining, and demonstrating effects among SUMSM.

Challenges associated with SUMSM provide a clear example of when behavioral interventions are necessary, but not sufficient. SUMSM are particularly difficult to engage due to factors directly associated with active substance use (i.e. homelessness, economic insecurity, sex work, psychological issues). SUMSM generally have higher baseline risk and more complex risk situations that make demonstrating sexual behavior change an elusive goal for current interventions. Behavioral interventions designed for SUMSM may benefit more from focusing on directly reducing substance use (e.g. contingency management or personalized cognitive counseling) or decreasing sexual behavior during substance use (Higa et al, 2013; Ostrow & Stall, 2008; Sullivan et al, 2012). In a combination prevention approach, it is likely that very high-risk SUMSM will require resource-intensive, individual-level interventions that address syndemic factors (i.e. childhood sexual abuse, mental health, homelessness) combined with biomedical interventions of pre and post-exposure prophylaxis (Chesney et al., 2003; Higa et al., 2013).

Peer-Led Interventions

Peer-led interventions were found to be more effective than other interventions (*OR*=.676, *p*=.002). This finding is consistent with Higa's 2013 review of EBIs; they

found that six out of nine EBIs were delivered by a peer facilitator. This study found seven out of 11 EBIs for HIV-negative MSM to be peer-led. Peers may aid in recruitment and retention, especially for high-risk or socially marginalized MSM who may not otherwise present for HIV prevention services. Peers may be perceived to be more credible and trustworthy than other professional facilitators, and can help increase retention by creating environments that feel non-judgmental, safe, and comfortable (Higa et al., 2013; Ye et al., 2014). A meta-analysis (Ye et al., 2014) of 15 peer-led interventions to reduce UAI among MSM showed that peer-led interventions decreased UAI overall, but effects varied by design type with the weakest effects being observed from RCTs. The evidence supporting peer-led interventions is very promising, especially for young MSM and MSM of color. More rigorous research is needed to understand factors related to effective peer-led interventions, and how peer-led components can best support combination approaches.

HIV Status

Trials were less effective when they had more HIV-positive participants *(p*=.02). The impact of HIV status was examined in three ways. First, meta-regression analysis of trials with any HIV-positive individuals showed that the effect size decreased with larger proportions of HIV-positive participants (Figure 10). Second, HIV status was dichotomized to indicate if the trial sample consisted of 20 percent or more HIV-positive MSM. These results were consistent with meta-regression; trials with less than 20 percent had stronger effects (*p*=.032). Finally, 11 trials excluded HIV-positive participants in their eligibility criteria. In subgroup

analyses, trials that excluded HIV-positive MSM had larger effects (*OR*=.765, *Q*(2)=9.742, *p*=.008).

Interventions designed for HIV-positive persons are a key component of CDC PRS's new prioritization agenda (CDC, 2015). Eligibility criteria for this study excluded trials designed exclusively for HIV-positive individuals because those trials were considered out of scope. However, most trials in this review allowed some proportion of HIV-positive individuals to enroll. Inclusion of HIV-positive participants may obscure the intervention's effect on HIV-negative MSM and contribute to challenges to demonstrate efficacy. This study's findings are consistent with Higa et al. (2013). Their review found that EBIs were more likely to exclusively focus on HIV-negative MSM. Currently, nine of 20 EBIs for MSM are recommended for HIV-positive MSM, yet not all were designed specifically for MSM. Previous meta-analysis suggests that HIV-positive individuals may benefit more from individual-level interventions rather than group-based, as well as interventions delivered in a clinical setting where they access other services (Crepaz et al., 2006). Similar to SUMSM, HIV-positive MSM represent a risk group where behavioral interventions are necessary, but not sufficient. Combination approaches that address medication adherence, biomedical options for sexual partners, and structural factors related to syndemics are necessary for HIV-positive MSM at high risk of transmitting to partners.

Education Level

Lower education level was found to be associated with effect size. Interventions with higher proportions of MSM with a high school education or less were shown to decrease overall effect size in meta-regression analyses (Figure 9) (p=.023). Reasons for this finding are not totally clear, but as stated previously, Higa found that trials with high methodological quality, but non-significant findings (i.e., "rigorous non-EBIs), were more likely to include MSM with lower education levels, as well as more substance users. Education is highly correlated with socioeconomic status and may indicate syndemics such as social or economic inequities (e.g. food insecurity, homelessness) that contribute to ongoing risk (van den Berg et al., 2015).

Intervention Level

Community-level interventions (CLIs) were found to be associated with greater effects (OR=.631, p=.032). Community-level interventions are different than individual or group-level interventions because they aim to change social norms and then measure behavior change at the community-level. CLIs are typically better-resourced and able to provide longer-term, multi-level interventions. This study's finding is consistent previous reviews demonstrating stronger effects among CLIs (Herbst et al., 2007; Johnson et al., 2008). It is possible that CLIs may be more effective than other interventions. CLIs may have more potential to address syndemic risk, and they may also be able to reach more MSM. Despite their potential reach, CLIs are usually a low dose intervention on a one-on-one level (i.e. brief encounters, handing out condoms) and unlikely to sustain individual behavior change. More

research is needed to answer questions. In this review, there were only three community-level trials (Kegeles et al., 2006; Kelly et al., 1991; Kelly et al., 1997), and one trial was a replication study (Kelly et al., 1997).

There was little difference in effect between individual and group-level interventions. In this study, more trials used group-level interventions (k=17) than individual-level (k=14). There is some research that suggests group-level interventions may have unintended negative outcomes for higher-risk MSM subpopulations because high-risk behaviors can be reinforced by peers met in small-group interventions (Johnson et al., 2005). However, evidence describing this effect is mixed. Crepaz (2006) and Herbst (2007) found individual-level interventions to offer the most benefit, especially for higher-risk populations requiring more intensive, tailored interventions such as substance users and HIV-positive MSM. Ye (2014) found peer-based interventions to only be effective in group-based interventions.

Retention

Trials with <80% retention showed slightly better effects (OR=.775, p=.09) than trials with better retention. Small sample size and poor retention were common reasons that trials did not meet criteria for methodological quality. High quality behavioral interventions targeting higher risk MSM such as substance-using MSM are likely to experience higher loss to follow up. Additionally, 50 percent of trials in this study reported use of an intention-to-treat (ITT) design; ITT designs are more likely to underestimate the effect when attrition is high and may further challenge demonstrations of efficacy. Finally, most trials used follow-up periods of more than 3

months (k=23), and 16 trials used follow-up periods of 12 months or more. However, median intervention time span was only three weeks. Follow-up periods that extend well beyond intervention termination are unlikely to demonstrate efficacy, especially among MSM subgroups that experience structural-level stressors, substance use, or mental health issues.

Implications

Despite a smaller than expected effect size, these results indicate that behavioral interventions continue to show some benefit for MSM. It is generally accepted that behavioral interventions are too small to decrease HIV incidence, but they can effectively reduce sexual risk behaviors over the short-term. Results from this study show that the effect size declined over time, challenging what is known about the relevance of behavioral interventions to reduce sexual risk behaviors. Over time, the effect size has gradually decreased and suggests that these interventions may soon be outdated if they are not improved.

To our knowledge, this is the first study that has examined effect size shrinkage of HIV behavioral interventions for MSM using cumulative meta-analysis. These summary findings lend strong support to arguments that behavioral interventions (and behavioral science) need to do better to improve effectiveness in the current social and political context. Further research, and future funding, is required to explain the most important reasons for effect size shrinkage, conduct replication studies to retire outdated components, and identify the most relevant and promising intervention features to use in combination approaches.

Qualitative review of presumably high-quality RCTs revealed lower methodological quality than expected and highlights limitations of the current evidence base. These findings are consistent with findings from CDC's Prevention Research Synthesis team (Higa et al., 2013). Despite 25 years of prevention research with MSM, there are only 11 out of 93 EBIs in the entire Compendium for HIV-negative MSM (an additional nine are for HIV-positive MSM) even though MSM are the most epidemiologically important risk group. Higa et al. (2013) described methodological challenges to demonstrating efficacy among MSM, but also pointed to misaligned funding priorities that result in the underfunding of high-quality research for MSM. Trials published after 2010 were less likely to have reporting biases or fatal flaws, and trials published after 2010 make up about half of trials eligible for this review. However, more than half of eligible trials were published prior to 2010 and influenced the overall effect size. In addition to efforts to make interventions work better, the research community would benefit from HIV prevention research standards that require high methodological quality (e.g., *a priori* power analyses, minimal retention requirements in both arms, criteria for missing data, replication studies, and required trial registration), standardized comparison across studies (e.g., similar outcome measures, standard control groups), and transparent reporting of results to facilitate high-quality research syntheses (e.g., CONSORT requirements; raw means and values).

This review found that trials using "integrated interventions" increased since 2008. Results suggest that intervening on and measuring diverse outcomes is feasible

and relevant for HIV-negative MSM. Future primary trials that use integrated approaches may be more efficient for HIV prevention because they can target multiple behavioral drivers of HIV risk among MSM. Integrated behavioral interventions used in combination with the most relevant biomedical strategies (e.g., interventions that decrease substance use and sexual risk behaviors with the additive benefit of PrEP for substance-using MSM) have high potential to curb HIV infection among high-risk MSM subgroups. Further research is needed to evaluate the effects of these approaches in real-world settings.

More research is needed to better understand how to use HIV behavioral interventions that support combination prevention approaches. A primary aim of this study was to re-evaluate HIV behavioral interventions in the context of new recommendations for combination approaches and identify the most promising factors associated with effectiveness. This was not possible due to unexpected statistical homogeneity. Combination prevention approaches are expected to focus intensive efforts on HIV subpopulations most at risk for HIV. Future research syntheses may benefit from narrowing the scope of reviews to focus on only these MSM subpopulations, and excluding MSM included in earlier trials that are no longer prioritized for HIV prevention interventions (i.e., white, older MSM of low to moderate HIV risk).

This review found that some priority MSM subpopulations were under-represented in current HIV research. Most notably, young MSM comprise the majority of new infections among MSM, but only accounted for 12 percent of eligible

trials. Six trials, or only 18 percent, were designed for substance-using MSM; nine trials, or 26 percent, were designed for MSM of color; and only one trial was specifically designed for MSM who are behaviorally bisexual (also have sex with women). There were no interventions specifically designed for emerging MSM subpopulations such as male sex workers or transgender women (who are sometimes classified under MSM). There were no RCTs designed for couple-level interventions. These findings highlight important research gaps. The research base is limited in its ability to examine how behavioral interventions would benefit emerging and priority subpopulations. Future primary trials need to prioritize the highest risk MSM. Research standards that set minimum risk requirements for inclusion should be established to ensure only high-risk MSM are enrolled in prevention trials.

Limitations

Findings must be interpreted within the context of this study's limitations. First, this study used a protocol that was very narrow in scope, and findings may not be perfectly comparable to other meta-analyses that included HIV-positive MSM, adolescent MSM, or MSM outside of the United States. Variation in study protocols and subjective decision-making throughout the research synthesis process can weaken the validity of comparisons between reviews and lead to spurious results (Ekkekakis, 2015). Future reviews in HIV prevention research should be guided by a standardized protocol so that the evidence for behavioral interventions can be regularly updated and compared to past reviews.

Second, a systematic review and meta-analysis cannot compensate for trials of low methodological quality. Trials published after 2010 showed higher quality than earlier trials, most likely resulting from trends in the research field such as the CONSORT statement (CONSORT, 2010). Earlier trials had more methodological flaws and higher reporting bias. Most early trials had some level of missing data or did not clearly report outcomes, statistical analyses, basic statistical information to calculate effect sizes (e.g., sample sizes by study arm; standard deviations), ITT, demographic information, or study procedures. This study excluded trials with missing data after attempts were made to contact study authors, resulting in a reduced set of trials for review. Additionally, many trials did not report *a priori* power analyses, and it is possible that many primary trials were underpowered. Future trials should aim to improve transparency in reporting to facilitate future evaluations of prevention research.

Third, sexual risk reduction outcomes are assessed through self-report and may be vulnerable to social desirability bias or recall bias. Many trials employed methods to reduce reporting bias such as using computer-assisted assessments or validated measures such as the Timeline Follow-Back. However, some trials did not clearly report how the outcome was measured or what procedures were used to increase confidentiality so as to reduce self-report bias.

Fourth, this study does not represent all MSM in the US. Important MSM subgroups were under-represented or missing from this review. For example, there were a surprisingly limited number of studies for young MSM or few studies specific

to MSM who also have sex with women. There were no studies specific to male sex workers, and very few trials that even reported the prevalence of male sex work. Further research on priority subgroups is essential to target behavioral interventions to the groups at highest epidemiological risk.

Fifth, the independent coder was only able to complete 50 percent of trial records. While there were few discrepancies, and all were able to be resolved through consensus, this still suggests threats to validity. Prior to publication, 100 percent of records will need to be reviewed by at least two coders.

Finally, this study's largest limitation was its high homogeneity. Lack of heterogeneity restricted this study from accomplishing its main goal of identifying the most promising and relevant intervention features of behavioral interventions for next generation combination approaches. While this study can offer an updated effect size and estimate of heterogeneity, findings from post-hoc moderator analyses must be interpreted with caution.

Conclusion

This study found evidence that the overall effect size for MSM-specific HIV behavioral interventions is still statistically significant. However, the magnitude of the effect size is considerably smaller than previous reviews, and if this trend continues, will slowly approach the null value as new trials are added. Gradual effect size decline over time suggests that behavioral interventions are becoming less effective, and must evolve to meaningfully contribute to HIV prevention in the current epidemic.

This study did not find evidence to suggest reliable moderators of effectiveness. Several variables were found to be potential moderators of effect, but will require additional research due to their post-hoc nature. Future research would benefit from further exploring if behavioral interventions are more effective when they are peer-led or when they have higher proportions of younger MSM. This study found promising results for the use of integrated behavioral interventions that address more than one problem behavior. Results suggest that integrated interventions have become more common in HIV prevention research, especially for high-risk MSM subgroups (i.e. substance-using MSM), although further research is needed to better understand how to improve effectiveness.

In conclusion, behavioral interventions remain necessary for HIV prevention among MSM, but are no longer sufficient. Behavioral interventions do not produce substantial effects to reduce HIV incidence among MSM, and effect size shrinkage raises questions about their continued ability to effectively reduce sexual risk behavior. Behavioral interventions may perform better when used in the context of new combination approaches, yet further research beyond this study is needed.

To improve behavioral interventions, future research (primary trials and systematic evaluations) should focus limited research resources on MSM subpopulations at highest risk of acquiring HIV (i.e., substance users, young MSM, MSM of color), and aim to identify the most relevant intervention components for the current epidemic. Replication studies of trials previously identified as effective interventions are needed to identify intervention components that are outdated and

need to be retired. Frequent, ongoing systematic reviews and meta-analyses are required to monitor the effectiveness of new approaches. Future research syntheses in this field should adopt routine cumulative meta-analysis to better monitor changes in the effect size as new evidence is accumulated.

APPENDIX A

EFFECTS OF HIV BEHAVIORAL INTERVENTIONS FOR MEN WHO HAVE SEX WITH MEN - UNITED STATES, 1988-2014

Systematic Review Search Strategy

Inclusion Criteria

Participants: Adult (<18) US men who have sex with men
Interventions: Any intervention with at least one behavioral sexual risk reduction component that states HIV prevention as a goal
Comparison: Any control (no treatment, usual care, attention control, another treatment)
Outcomes: Behavioral and biological outcomes.
Study design: Randomized controlled trials in which participants were prospectively assigned to study groups and in which control group outcomes were measured concurrently with intervention group outcomes.

Exclusion Criteria

Participants: Individuals who live outside the US, adolescents <18, 100% HIV-positive samples
Study design: Any study without a contemporaneous control group

Filter: No filters except date (January 1, 1988 – current date)

Search String 1. MSM terms (population filter): MSM, men who have sex with men, gay, homosexual, bisexual, transgender

Search String 2. HIV terms (disease filter): HIV, AIDS, STD (MeSH subheading prevention & control when available)

Search String 3. Cochrane published RCT filter (study design filter): (randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans [mh])

Databases

1. PubMed
2. EMBASE
3. CENTRAL
4. PsycINFO
5. CINAHL

PUBMED

Searched 04/26/15

Limits: Publication Date 1988-2015

Notes: Export by batch to Endnote (50 at a time); Delete “search” and post-script, filters, sort by, etc. when re-running. Sort by Pub Date. Export 500 per time to EndNote.

Search	Query	Items found
#10	Search ((#1) AND #3) AND #7	1208
#9	Search ((#1) AND #3) AND #7 Filters: Publication date from 1988/01/01 to 2015/12/31 Sort by: PublicationDate	1208
#8	Search (((randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans [mh])))	3064779
#7	Search (((randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR placebo [tiab] OR drug therapy [sh] OR randomly [tiab] OR trial [tiab] OR groups [tiab]) NOT (animals [mh] NOT humans [mh]))) Filters: Publication date from 1988/01/01 to 2015/12/31 Sort by: PublicationDate	2554293
#6	Search (#1) AND #3	4361
#5	Search (#1) AND #3 Filters: Publication date from 1988/01/01 to 2015/12/31 Sort by: PublicationDate	4361
#4	Search ((((((((((“HIV Infections/prevention and control"[Mesh] OR "Acquired Immunodeficiency Syndrome/prevention and control"[MeSH Terms] OR "Sexually Transmitted Diseases/prevention and control"[MeSH Terms]))))))))	50184
#3	Search ((((((((((“HIV Infections/prevention and control"[Mesh] OR "Acquired Immunodeficiency Syndrome/prevention and control"[MeSH Terms] OR "Sexually Transmitted Diseases/prevention and control"[MeSH Terms])))))))) Filters: Publication date from 1988/01/01 to 2015/12/31 Sort by: PublicationDate	46377

#2	Search (((((((("men"[MeSH Terms] OR "men"[All Fields]) AND who[All Fields] AND ("sex"[MeSH Terms] OR "sex"[All Fields]) AND ("men"[MeSH Terms] OR "men"[All Fields])) OR "msm"[All Fields])) OR "men who have sex with men"[All Fields]) OR ("homosexuality, male"[MeSH Terms] OR ("homosexuality"[All Fields] AND "male"[All Fields]) OR "male homosexuality"[All Fields] OR "gay"[All Fields])) OR ("bisexuality"[MeSH Terms] OR "bisexuality"[All Fields] OR "bisexual"[All Fields])) OR ("homosexuality"[MeSH Terms] OR "homosexuality"[All Fields] OR "homosexual"[All Fields])) OR ("transgendered persons"[MeSH Terms] OR ("transgendered"[All Fields] AND "persons"[All Fields]) OR "transgendered persons"[All Fields] OR "transgender"[All Fields])))	61967
#1	Search (((((((("men"[MeSH Terms] OR "men"[All Fields]) AND who[All Fields] AND ("sex"[MeSH Terms] OR "sex"[All Fields]) AND ("men"[MeSH Terms] OR "men"[All Fields])) OR "msm"[All Fields])) OR "men who have sex with men"[All Fields]) OR ("homosexuality, male"[MeSH Terms] OR ("homosexuality"[All Fields] AND "male"[All Fields]) OR "male homosexuality"[All Fields] OR "gay"[All Fields])) OR ("bisexuality"[MeSH Terms] OR "bisexuality"[All Fields] OR "bisexual"[All Fields])) OR ("homosexuality"[MeSH Terms] OR "homosexuality"[All Fields] OR "homosexual"[All Fields])) OR ("transgendered persons"[MeSH Terms] OR ("transgendered"[All Fields] AND "persons"[All Fields]) OR "transgendered persons"[All Fields] OR "transgender"[All Fields]))) Filters: Publication date from 1988/01/01 to 2015/12/31 Sort by: PublicationDate	52369

EMBASE

Searched 04/26/15
Limits: Publication Date 1988-2015
Notes: Sorts by Pub Year by default; export records to CSV and RIS (EndNote); provides detailed queries upon export that specify contents of combined searches (i.e. #1 AND #3; #6 is complete final string)

No.	Query EMBASE 04/26/15	Results
#6	'men who have sex with men'/exp OR 'men who have sex with men' OR 'homosexual male'/exp OR 'homosexual male' OR 'homosexuality'/exp OR 'homosexuality' OR 'male homosexuality'/exp OR 'male homosexuality' OR 'bisexual male'/exp OR 'bisexual male' OR 'bisexuality'/exp OR 'bisexuality' OR 'transgender'/exp OR 'transgender' OR 'msm' OR 'gay' AND ('sexually transmitted diseases'/exp/dm_pc OR 'acquired immunodeficiency syndrome'/exp/dm_pc OR 'human immunodeficiency virus infection'/exp/dm_pc) AND ('randomized controlled trial'/exp OR 'randomized controlled trial' OR 'controlled clinical trial'/exp OR 'controlled clinical trial' OR randomized:ti OR randomized:ab OR placebo:it OR placebo:ab OR randomly:it OR randomly:ab OR groups:it OR groups:ab OR trial:it OR trial:ab OR 'drug therapy'/de OR 'intervention study'/de) NOT ('animals'/exp OR animals NOT ('humans'/exp OR humans)) AND (1988:py OR 1989:py OR 1990:py OR 1991:py OR 1992:py OR 1993:py OR 1994:py OR 1995:py OR 1996:py OR 1997:py OR 1998:py OR 1999:py OR 2000:py OR 2001:py OR 2002:py OR 2003:py OR 2004:py OR 2005:py OR 2006:py OR 2007:py OR 2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py)	**766**
#5	'men who have sex with men'/exp OR 'men who have sex with men' OR 'homosexual male'/exp OR 'homosexual male' OR 'homosexuality'/exp OR 'homosexuality' OR 'male homosexuality'/exp OR 'male homosexuality' OR 'bisexual male'/exp OR 'bisexual male' OR 'bisexuality'/exp OR 'bisexuality' OR 'transgender'/exp OR 'transgender' OR 'msm' OR 'gay' AND ('sexually transmitted diseases'/exp/dm_pc OR 'acquired immunodeficiency syndrome'/exp/dm_pc OR 'human immunodeficiency virus infection'/exp/dm_pc) AND ('randomized controlled trial'/exp OR 'randomized controlled trial' OR 'controlled clinical trial'/exp OR 'controlled clinical trial' OR randomized:ti OR randomized:ab OR placebo:it OR placebo:ab OR randomly:it OR randomly:ab OR groups:it OR groups:ab OR trial:it OR trial:ab OR 'drug therapy'/de OR 'intervention study'/de) NOT ('animals'/exp OR animals NOT ('humans'/exp OR humans))	773

#4	'randomized controlled trial'/exp OR 'randomized controlled trial' OR 'controlled clinical trial'/exp OR 'controlled clinical trial' OR randomized:ti OR randomized:ab OR placebo:it OR placebo:ab OR randomly:it OR randomly:ab OR groups:it OR groups:ab OR trial:it OR trial:ab OR 'drug therapy'/de OR 'intervention study'/de NOT ('animals'/exp OR animals NOT ('humans'/exp OR humans))	2734764
#3	'men who have sex with men'/exp OR 'men who have sex with men' OR 'homosexual male'/exp OR 'homosexual male' OR 'homosexuality'/exp OR 'homosexuality' OR 'male homosexuality'/exp OR 'male homosexuality' OR 'bisexual male'/exp OR 'bisexual male' OR 'bisexuality'/exp OR 'bisexuality' OR 'transgender'/exp OR 'transgender' OR 'msm' OR 'gay' AND ('sexually transmitted diseases'/exp/dm_pc OR 'acquired immunodeficiency syndrome'/exp/dm_pc OR 'human immunodeficiency virus infection'/exp/dm_pc)	3648
#2	'sexually transmitted diseases'/exp/dm_pc OR 'acquired immunodeficiency syndrome'/exp/dm_pc OR 'human immunodeficiency virus infection'/exp/dm_pc	47878
#1	'men who have sex with men'/exp OR 'men who have sex with men' OR 'homosexual male'/exp OR 'homosexual male' OR 'homosexuality'/exp OR 'homosexuality' OR 'male homosexuality'/exp OR 'male homosexuality' OR 'bisexual male'/exp OR 'bisexual male' OR 'bisexuality'/exp OR 'bisexuality' OR 'transgender'/exp OR 'transgender' OR 'msm' OR 'gay'	49275

CENTRAL

Searched 04/26/15

Limits: Publication Year 1988-2015, Clinical Trials

Notes: Do not use RCT filter; use MeSH terms when possible; Export to txt file with title and abstract

1	MeSH descriptor: [HIV Infections] explode all trees and with qualifier(s): [Prevention & control - PC]	1944
2	MeSH descriptor: [Acquired Immunodeficiency Syndrome] explode all trees and with qualifier(s): [Prevention & control - PC]	228
3	MeSH descriptor: [Sexually Transmitted Diseases] explode all trees and with qualifier(s): [Prevention & control - PC]	2172
4	#1 or #2 or #3 Publication Year from 1988 to 2015, in Trials	1875
5	"MEN WHO HAVE SEX WITH MEN" or "MSM"	447
6	MeSH descriptor: [Homosexuality, Male] explode all trees	233
7	HOMOSEXUALITY and MALE	373
8	"MALE HOMOSEXUALITY"	18
9	MeSH descriptor: [Homosexuality] explode all trees	338
10	HOMOSEXUALITY or HOMOSEXUAL	543
11	MeSH descriptor: [Bisexuality] explode all trees	43
12	BISEXUALITY or BISEXUAL	133
13	MeSH descriptor: [Transgendered Persons] explode all trees	2
14	"TRANSGENDERED" and "PERSONS" or "TRANSGENDERED PERSONS" or TRANSGENDER	36
15	GAY	469
16	#5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 Publication Year from 1988 to 2015, in Trials	827
17	#4 and #16 Publication Year from 1988 to 2015, in Trials	116

PSYCINFO

Searched 04/26/15

Limits: 1988-2015

Notes: Unable to use MeSH terms specific to DISEASE prevention & control resulting in a more sensitive string with more hits; view all in abstract view, 50 per page, and then save to folder 50 at a time. Export entire folder to EndNote. Save hyperlink with search string.

S1	"men who have sex with men" OR "msm" OR "homosexual" OR "male homosexuality" OR "homosexuality" OR ("homosexual" AND "male") OR "gay" OR "gay men" OR ("gay" AND "men") OR "bisexual" OR ("bisexual" AND "men") OR "bisexuality" OR "transgender persons" OR ("transgender" AND "persons") OR "transgender" OR (((ZU "homosexuality")) or ((ZU "bisexuality"))) or ((ZU "transgender"))	27,805
S2	(((ZU "hiv")) or ((ZU "aids") or (ZU "aids prevention"))) or ((ZU "sexually transmitted diseases"))	38,015
S3	(PT randomized controlled trials OR PT controlled clinical trials OR TI randomized OR AB randomized OR TI placebo OR AB placebo OR AB trial OR TI randomly OR AB randomly OR TI groups OR AB groups OR TI trial) OR DE drug therapy NOT ((animals NOT humans))	877,947
S4	#S1 AND #S2	5,794
S5	#S4 AND #S3	1,803

CINAHL

Searched 04/27/15

Limits: 1988-2015

Notes: Unable to use MeSH terms specific to DISEASE prevention & control resulting in a more sensitive string with more hits; view all in abstract view, 50 per page, and then save to folder 50 at a time. Export entire folder to EndNote. Same platform as psycINFO.

S1	("men who have sex with men" OR "msm" OR "homosexual" OR "male homosexuality" OR "homosexuality" OR ("homosexual" AND "male") OR "gay" OR "gay men" OR ("gay" AND "men") OR "bisexual" OR ("bisexual" AND "men") OR "bisexuality" OR "transgender persons" OR ("transgender" AND "persons") OR "transgender") OR ((MH "Homosexuality+") OR (MH "Homosexuals+") OR (MH "Homosexuals, Male+")) OR ((MH "Bisexuality+") OR (MH "Bisexuals+ ")) OR (MH "Transgendered Persons+")	9,247
S2	(MH "Human Immunodeficiency Virus+") OR (MH "Acquired Immunodeficiency Syndrome/PC") OR (MH "Sexually Transmitted Diseases+/PC")	19,371
S3	#S1 AND# S2	1,523
S4	(MH "Clinical Trials") OR (MH "Community Trials") OR (MH "Randomized Controlled Trials") OR (MH "Preventive Trials") OR (MH "Intervention Trials") OR TI randomized OR AB randomized OR TI placebo OR AB placebo OR AB trial OR TI randomly OR AB randomly OR TI groups OR AB groups OR TI trial OR DE drug therapy NOT ((animals NOT humans))	292,348
S5	#S3 AND #S4	245

APPENDIX B: STUDY PROTOCOL

Study ID# (add report ID# if more than 1 report from same study)	Use pre-specified study ID# from coding assignment tab (example: 01). If a study is identified as having multiple reports, use a report ID# (Example 01-A; use a letter).
Date of Review	The date you conducted the review (e.g. 08/07/15)
Coder Initials	Your initials (e.g. EL - Elsa Larson)
Author Last Name	Author's last name for reference
Year	Year of publication
Key Words	Provide 3-5 key words that you think are important to describe this article (i.e. bisexual men, methampethamine-users, internet-based intervention, young black MSM, etc).
Notes (if any)	Notes/concerns/questions for this article
ELIGIBILITY REVIEW. IF NO TO ANY QUESTION=NOT ELIGIBLE. STOP REVIEW.	The following questions are the eligibility criteria for the review. If NO to any question, STOP review. Note reason why.
Is this study published in a peer-reviewed journal? (yes/no)	Identify journal name and confirm it uses a peer-review process
Was this study specifically designed for MSM (y/n)	Did the authors describe they designed this for MSM, and tested it on MSM?
Are adult US MSM the primary study population? (y/n)	Are U.S. MSM over 18 the primary study population, >95%? MSM may be of any race/ethnicity, sexual identity, or HIV-status, etc. Exclude studies that focus on non-US MSM, adolescents (<18). If mean age is 18 or older, include.

Is this study a behavioral intervention? (y/n)	Scan article to identify if the study is categorized as a Behavioral intervention. Behavioral Interventions try to change individual risk behaviors by providing necessary skills or materials. Intervention methods might involve modeling or demonstration, role-playing, or participatory skill development (i.e. increasing condom use). Behavioral interventions are different than provision-of-information-only interventions that try to change knowledge, attitudes, or norms only (e.g. increasing HIV knowledge), or environmental interventions that aim to change the physical or social environment to promote health and prevent disease (e.g. reducing stigma). Exclude information-only, environmental, and pharmaceutical interventions (i.e. PreP)
Was this study tested in RCT with an independent comparison group? (y/n)	Include only randomized controlled trials with an independent comparison group. (i.e. treatment and control, pre and post measurements for each group).
Did this study measure HIV behavioral or biological outcomes? (y/n)	Include studies that describe HIV prevention, or behavioral risk reduction, as the primary or secondary study outcome. HIV prevention outcomes may be behavioral (e.g. increasing condom use, decreasing # of sexual partners, etc) or biological (e.g. HIV/STI incidence or prevalence). Include studies that may have had a different primary outcome (e.g. alcohol or substance use behaviors), but included HIV prevention or behavioral risk reduction as a secondary outcome.
100% HIV positive sample?	If 100% of participants were HIV-positive, exclude.
Include or exclude?	Decision
If excluded, why?	If you excluded, indicate why the study was not eligible
NOTES	Any notes

Participant Characteristics	If any variable is unclear or not reported, mark UNCLEAR or NOT REPORTED
Sample size at baseline (enrollment)	Report N at baseline (enrollment). This may be what they report in the abstract
Age Range	Report the age range of participants included in the study (entire sample)
Age (mean)	Report the mean age for the entire sample at baseline
Age, SD	Report the age standard deviation
age (median)	Report the age median, if available
(%) race/ethnicity - (Black, Latino/Hispanic, white, other)	Report the % of white, black, Asian, Hispanic, etc, for the entire study sample. Studies will vary in how they report race and ethnicity. Report Black, Latino/Hispanic, White, and Other.
% racial/ethnic minority MSM	Calculate the sum of Hispanic and non-white MSM in the sample and report as % of minority race/ethnicity msm represented in the study
% HIV-Positive	Report % of sample that was HIV-positive
How was HIV-status assessed?	How did the authors measure HIV status (i.e. self-report, testing at baseline?)
(%) HS education or less	Report percent that graduated HS or less than HS.
Would you say this is a very high-risk MSM sample? (y/n)	In your opinion, do you think this is a high-risk sample as defined by high numbers of baseline sexual partners, risky sex and drug activity, high-risk sexual networks, etc. You do not need to qualify.
% substance users	Report % of study sample who were substance users (any)
% sex workers/trade sex	Report % of study sample who were sex workers/traded sex
% MSMW (not identified as bisexual)	Report % of study sample who had sex with both men and women, but did not identify as bisexual
% gay-identified	Report % of study sample who identified as gay
% bisexual-identified	Report % of study sample who identified as bisexual

% mental health or vicitimization history	Report % of study sample who had mental health issues, including but not limited to depression, childhood sexual abuse, lifetime victimization, mental distress other.
Intervention Characteristics	
Any biological outcomes?	Report if the study used biological outcomes as endpoints
Intervention level (individual-level, group-level, community-level)	Report if the study was an individual-level intervention (ILI), group-level intervention (GLI), or community-level intervention (CLI) - as stated in article - if unclear, mark unclear
Randomized?	Were participants randomized to study groups? (yes, no). If not, report how they were assigned to arms.
N randomized?	Report N they randomized
Method of randomization/allocation	How did the authors operationalize the randomization procedure?
Unit of randomization	individual, group, community?
Unit of analysis	individual, group, community?
If unit of randomization does not match unit of analysis, did authors control for this? E.g. report intra-class correlation coefficient?	Report only if unit of randomization and unit of assignment are different. If they are different, did authors report that they controlled for it (e.g. intra-class correlation coefficient)?
More than 1 experimental condition?	Note yes/no if study used more than 1 experimental condition (e.g. two group randomized trial with control)
comparison/control condition (description of control)	Describe the control group in 1-2 words (information only, usual treatment, general health education, etc)
Classification of control (wait list control, HIV-related comparison group, non-HIV comparison group)	categorize as wait-list, HIV-related comparison (any HIV information, etc), non-HIV related (i.e. nutrition workshop)
Intervention setting (Community, clinic, MSM setting, other)	Where did the intervention take place?

Who facilitated the intervention? (MSM, professional, paraprofessional)	Who facilitated/delivered the intervention? Was it another MSM, professional (e.g. mental health counselor), para-professional (e.g. HIV test counselor, community outreach worker)?
Peer-led intervention?	Indicate if it was a peer-led intervention. It may not be explicit, so examine if the known facilitator matches the participants, if so, then classify as peer-led.
Group size? (if GLI)	If GLI - about how many participants in each group?
Intervention time span (# of weeks from start to finish)	How many weeks did the intervention last from start to finish? (e.g. 6 weeks, 12 weeks, etc)
Number of sessions	How many total sessions were in the intervention (e.g. 2, 6, 12)
Number of hours (total time of intervention)	How many hours (or total time) was the intervention? (e.g. 3 sessions X 2 hours each = 6 hours total dose)
Last recall period for primary outcome	What the recall period participants were asked to use for the last follow-up session? (e.g. ppts asked to report on last **30 days** of sexual behavior=30 days)
Incentives for participation?	Did participants receive incentives for participation? (yes/no)
Method of outcome assessment (interview, paper survey, ACASI, internet survey)	How were the outcomes assessed? Use categories: interview, paper survey, ACASI/CASI, internet survey)
Years conducted	When was the study started and finished? (e.g. 1997-1998)
Multi-site or single site?	If interventions were conducted at one site, list as single-site. If interventions were conducted across multiple sites, list as multi-site.
US region	Report US region if known
Intervention Name/Title	What did the authors call the intervention? (i.e. Project RESPECT, Many Men Many Voices - 3MV, Living Well, etc…). If there is no name, report NR.

Theory	Did the authors report a theoretic principle (e.g. Transtheoretical Model/Stages of Change, Information-Motivation-Behavior, Social Cognitive Theory, Motivational Interviewing, etc)? If so, list.
Did the authors report it was pilot tested?	Did the authors describe any pilot testing or exploratory research prior to intervention launch?
Include more than 1 behavioral outcome?	Does the study measure outcomes on at least 1 other behavior (other than sexual risk reduction; i.e. substance use, HIV testing)?
Focus on racial/ethnic minority MSM?	Did the authors describe any known psychometric properties of the scale (i.e. previously tested, cronbach's alpha/reliability coefficients, etc).
Focus on MSM substance users?	yes/no
Other specified subgroup?	Was this intervention designed, tailored to, or otherwise focused on MSM of color? (yes/no) If yes, list specific group
Focus on technology? (if yes, describe)	Did this intervention focus on using technology such as the Internet, telephones, text messaging, or other communication channels?
Focus on sexual communication? (yes/no)	Did this intervention focus on improving or increasing any kind of sexual communication between partners (e.g. HIV status disclosure, condoms, safer sex, etc?
Focus on individualized risk reduction plans? (yes/no)	Did the intervention components include development of individualized risk reduction plans tailored to the individual?
Focus on stigma and discrimination? (yes/no)	Did the intervention focus on stigma and discrimination (related to being gay, or being MSM of color) in intervention components?
Inclusion Criteria (HIV status, substance use, UAI, serodiscordant partner…)	list inclusion criteria as described exactly by the authors
Study Results	

Primary outcome (s)	What is the study's intended outcome? (e.g. increase condom use). If multiple outcomes, list all; separate with comma/semi-colon
Any significant positive effects on outcome/s? (difference between treatment and control)	Were the statistical results significant between groups and in the right direction? Yes/no (e.g. yes=study showed 26% decrease in UAI that was statistically significant)
If other interesting results, describe here (i.e. within groups differences, etc).	Use this spot to report on other significant or interesting results including within groups (pre-post change scores by group), subgroup analyses, etc.
Significance testing based on α=.05 (or less) and 2-sided test?	Were significance tests based on .05 alpha level (or more stringent such as .01) and a two-tailed test?
Aim: reductions in UAI? (list sig/ns)	If UAI (unprotected anal intercourse) was the outcome, were outcomes significant (sig) or not significant (ns)?
Aim: increased condom use? (sig/ns)	If condom use was the outcome, were outcomes significant or not significant?
Aim: reduce number of sexual partners?	if # of partners was an outcome, were outcomes significant or not significant?
Aim reduced substance use? (sig/ns)	If substance use was the outcome, were outcomes significant or not significant?
Aim: reduce serodiscordant sex? (sig/ns)	If serodiscordant sex was the outcome, were outcomes significant or not significant?
Aim: reduce receptive sex? (sig/ns)	If receptive sex was the outcome, were outcomes significant or not significant?
Other outcomes (testing, disclosure, substance use, list sig/ns)	If other outcomes were studies such as HIV testing, disclosure to partners, or substance use, describe and report if results were sig or ns.
A priori power analysis conducted?	Did the authors report that they conducted an a priori power analysis to specify sample size? If so, did they specify >80% power?
Follow-up time used for analyses (# months after T1)	How many months between T1 and last follow up (T2)?
Analytic sample >40 per group?	Did the treatment and control groups have at least 40 ppts in each at the time of analysis?
n treatment group at T1 (baseline)	How many ppts in the tx group at T1 (at baseline)?

n control group at T1	How many ppts in the control group at T1 (at baseline)?
n tx group T2 (follow-up)	How many ppts in the tx group at T2 (at follow-up)?
n control group T2	How many ppts in the control group at T2 (at follow-up)?
% and n - retention at last f/u	At the last follow-up, what percent of ppts overall were retained in the study? (compare overall baseline to overall f/u and calculate %)
% retention tx group (at last f/u)	At the last follow-up, what percent of ppts in the tx group were retained?
% retention in control group (at last f/u)	At the last follow-up, what percent of ppts in the control group were retained?
Differential Retention?	Is there differential retention between tx and control groups?
Baseline differences between study groups?	Were there baseline differences between the tx and control groups? Yes/no
If yes, what?	
(ITT) Intent-to-treat analysis (ppts analyzed in original groups and data replacement for dropouts)?	Did the authors specify they used an ITT analysis?
Did study provide a CONSORT statement?	Did the article have an attachment or reference to a CONSORT statement? (standardized reporting for RCTs)
Was study a registered trial?	Did the study indicate a trial registration # in the article or abstract?
Statistical Information	
For continuous data analyses (T1=baseline; T2=final f/u)	**Only for interval-level data - skip to categorical if needed - retrieve baseline (T1) and final follow/up (T2). Retrieve n for each of the four groups (baseline tx, baseline control, f/u tx, f/u control), the group means, and their SD or SE to calculate effect sizes later. If there are multiple outcomes (e.g. reduced UAI, reduced substance use), use multiple rows and list study ID# each time**

Study ID (use multiple lines for multiple outcomes)	ID
Specific outcome variable definition	Describe the outcome as specifically as possible, matching the authors description
N analyzed for results	Total N for analytic sample
N analyzed in tx group	n for tx group
N analyzed in control group	n for control group
Means transformed or adjusted? If so indicate with red type	Indicate if any means were transformed or adjusted with yes/no and then mark the reported adjusted means with a (*).
N analyzed at T1 - tx group	n for how many in tx group at time 1
Mean at T1 (tx group)	group mean for the tx group at T1
SD at T1 (tx group)	If reported, list standard deviation
SE at T1 (tx group)	if reported, list standard error
N analyzed at T1 - control group	n for how many in control group at time 1 (baseline)
Mean at T1 (control group)	group mean for the tx group at T2
SD at T1 (control group)	If reported, list standard deviation
SE at T1 (control group)	if reported, list standard error
N analyzed at T2 - tx group	n for how many in tx group at time 2 (last follow-up)
Mean at T2 (tx group)	What was the follow-up mean for the tx group? (e.g. mean number of UAI occasions at T1). Use last possible follow-up.
SD at T2 (tx group)	If reported, list standard deviation
SE at T2 (tx group)	if reported, list standard error
N analyzed at T2 - control group	n for how many in control group at time 2 (f/u)
Mean at T2 (control group)	What was the follow-up mean for the control group? (e.g. mean number of UAI occasions at T1). Use last possible follow-up.
SD at T2 (control group)	list standard deviation
SE at T2 (control group)	list standard error
Mean Difference Tx group (T1-T2)	calculate mean differences; use excel formula

Mean difference control group (T1-T2)	calculate mean differences; use excel formula
Mean Difference (T - C at T2)	calculate mean differences; use excel formula
correlation pre-post scores	if reported, otherwise leave blank
statistical method used	Report test used
t-value	t-value if reported
F statistic	F-stat if reported
p-value	p-value if reported
Df	df if reported
RR	RR if reported (e.g. ratio of mean # of events in Tx group to mean # of events in control, see Johnson's 2008 Cochrane review)
Beta	betas for any regression models
SMD	Standardized mean difference if reported. If unclear, consult Cochrane definition at: http://handbook.cochrane.org/chapter_9/9_2_3_2_the_standardized_mean_difference.htm
95% CI	If the authors reported 95% CI for ES, list
Other	
Effect size reported	Yes/no
ES type	type (Cohen's d, Cohen's f, etc)
ES	test statistic
95% CI for ES	95% CI specific to ES
For categorical data analysis	**for dichotomous data: Think about constructing a 2X2 table for each of the four groups; we need the # or % who reported the outcome ("yes" = ex. this would be the 1 group in a binary analysis) and the # or % that did NOT report the outcome ("no"= ex. this would be the 0 group in a binary analysis). If there are multiple outcomes (e.g. reduced UAI, reduced substance use), use multiple rows, and list study ID# each time.**
Specific outcome variable definition	specific outcome variable, same as continuous
N analyzed tx group	Instructions same as continuous

N analyzed control group	Instructions same as continuous
N analyzed at T1-tx group	Instructions same as continuous
number for tx group at T1 - outcome	If authors reported counts, list number of ppts in tx group that reported outcome at T1 (e.g. 50 report any UAI).
number for tx group at T1 - no outcome	If authors reported counts, list number of ppts in tx group that reported no outcome at T1 (e.g. 50 reported zero UAI).
proportion for tx group at T1 - outcome	If authors reported % or proportion, list number of ppts in tx group that reported any outcome at T1 (e.g. 50% reported UAI).
proportion for tx group at T1 - no outcome	If authors reported % or proportion, list number of ppts in tx group that reported no outcome at T1 (e.g. 50% reported NO UAI).
N analyzed at T1- control group	same as continuous
number for control group at T1 - outcome	If authors reported counts, list number of ppts in control group that reported outcome at T1 (e.g. 50 report any UAI).
number for control group at T1 - no outcome	If authors reported counts, list number of ppts in control group that reported no outcome at T1 (e.g. 50 reported zero UAI).
prop. for control group at T1 - outcome	If authors reported % or proportion, list % of ppts in control group that reported any outcome at T1 (e.g. 50% reported UAI).
prop. for control group at T1 - no outcome	If authors reported % or proportion, list % of ppts in control group that reported no outcome at T1 (e.g. 50% reported NO UAI).
N analyzed at T2-tx group	Instructions same as continuous
number for tx group at T2 - outcome	If authors reported counts, list number of ppts in tx group that reported outcome at T2 (e.g. 50 report any UAI).
number for tx group at T2 - no outcome	If authors reported counts, list number of ppts in tx group that reported no outcome at T2 (e.g. 50 reported zero UAI).
prop. for tx group at T2 - outcome	If authors reported % or proportion, list % of ppts in tx group that reported any outcome at T2 (e.g. 50% reported UAI).

prop. for tx group at T2 - no outcome	If authors reported % or proportion, list % of ppts in tx group that reported no outcome at T2 (e.g. 50% reported NO UAI).
N analyzed at T2-control group	same as continuous
number for control group at T2 - outcome	If authors reported counts, list number of ppts in control group that reported outcome at T2 (e.g. 50 report any UAI).
number for control group at T2 - no outcome	If authors reported counts, list number of ppts in control group that reported no outcome at T2 (e.g. 50 reported zero UAI).
prop. for control group at T2 - outcome	If authors reported % or proportion, list number of ppts in control group that reported any outcome at T2 (e.g. 50% reported UAI).
prop. for control group at T2 - no outcome	If authors reported % or proportion, list number of ppts in control group that reported no outcome at T2 (e.g. 50% reported NO UAI).
statistical method used	statistical test used
$\chi 2$	list test stats that are reported
OR	Odds ratio
RR	Relative risk
PR	Prevalence ratio
Risk Diff	Risk difference
CI	95% confidence interval
Logged OR	Logged Odds Ratio
p value	Exact p-value
Other	Other
effect size reported?	Yes/no
ES type	type (Cohen's d, Cohen's f, etc)
95% CI for ES	95% CI for ES
NOTES	Notes
Methodological Quality Review using PRS Criteria	Refer to PRS criteria http://www.cdc.gov/hiv/dhap/prb/prs/efficacy/rr/criteria/index.html
Is study described as an EBI?	Is it reported to be an Evidence-Based Intervention?
If CDC already classified in their 2013 review (studies before 2011), list their ranking	Use Higa et al (2013) when possible

(EBI, positive non-EBI, rigorous non-EBI, other non-EBI)	
Refer to PRS GOOD criterion for study's intervention level (ILI, GLI, CLI): Indicate if study met all points for each criterion; if no, list failures.	yes/no
I. Intervention Description	yes/no
II. Quality of Study Design	yes/no
III. Quality of Study Implementation and Analysis	yes/no
IV. Strength of Evidence	yes/no
V. No demonstrated significant negative intervention effects	yes/no
VI. Additional Limitations to Evaluate	yes/no
Methodological Quality Category	
EBI: **Met PRS criteria for good and had positive result**	yes/no
Rigorous Non-EBI: **Met good PRS criteria, but no positive result**	yes/no
Positive Non-EBI: **Positive Result, but did not meet at least 1 PRS criteria**	yes/no
Other-Non-EBI: **No positive result, did not meet at least 1 PRS criterion**	yes/no
Different from CDC rating, when applicable	Notes
NOTES	
DONE! STOP.	

APPENDIX C

COMPENDIUM OF EVIDENCE-BASED INTERVENTIONS AND BEST PRACTICES FOR HIV PREVENTION RISK REDUCTION (RR) CHAPTER
Last updated December 3, 2014

PRS Efficacy Criteria for Best-Evidence Risk Reduction (RR) Individual-level, Group-level, and Couple-level Interventions (ILIs/GLIs/CPLs)

Intervention Description
- Clear description of key aspects of the intervention

Quality of Study Design
- Prospective study design
- Appropriate and concurrent comparison arm
- Random or minimally biased assignment of subjects to study arms

Quality of Study Implementation and Analysis
- Follow-up assessment ≥ 3-months post completion of intervention for each study arm with recall not referring to pre-intervention period)
- At least a 70% retention rate at a single follow-up assessment for each study arm
- Comparison between intervention arm and an appropriate comparison arm
- Analysis of participants subjects in study arms as originally allocated regardless of contamination or logistic/implementation issues
- Analysis of participants regardless of the level of intervention exposure
- Use of appropriate cluster-level analyses if assigned to study arms by cluster or group
- Analysis must be based on post-intervention levels or on pre-post changes in measures
- For pre-post changes used in analysis, measures must be identical, including identical recall period
- Analysis based on an $\alpha = .05$ (or more stringent) and a 2-sided test
- With nonrandomized assignment, either no statistical differences in baseline levels of the outcome exist or baseline differences are controlled for in the analysis
- Analytic sample ≥ 50 participants per study arm

Strength of Evidence
Demonstrated Significant Positive Intervention Effects
- Positive and statistically significant ($p < .05$) intervention effect for ≥1 relevant outcome measure
- A positive intervention effect is defined as a greater reduction in HIV/STD incidence or risk behaviors or a greater increase in HIV protective behaviors in the intervention arm relative to the comparison arm
- A relevant outcome is defined as a behavior (e.g., abstinence, mutual monogamy, number of sex partners, consistent condom use with anal/vaginal sex, unprotected anal/vaginal sex, proportion of anal/vaginal sex acts protected, injection drug use, sharing or borrowing needles/works) - that directly impacts HIV risk or a biologic measure indicating HIV or STD infection (i.e., HIV or STD incidence)

- Effect at the follow-up and based on the analyses that meet study implementation and analysis criteria

No Demonstrated Significant Negative Intervention Effects

- No negative and statistically significant ($p < .05$) intervention effect for any relevant outcome
- A negative intervention effect is defined as a greater increase in HIV/STD incidence or risk behaviors or a greater decrease in HIV protective behaviors in the intervention arm relative to the comparison arm.
- No other statistically significant harmful intervention effect
- For an intervention with a replication evaluation, no significant negative intervention effects in the replication study

Additional Limitations to Evaluate

- No evidence that additional limitations resulted in a fatal flaw: o A fatal flaw has occurred when the overall evaluation of limitations indicates they resulted in considerable bias, thus substantially reducing the confidence of the findings.
- Examples of item limitations to check for possible fatal flaw: Effects only found within potentially biased subset analyses;
- Substantial missing data. Missing data plus loss to attrition exceeds acceptable limits for retention alone (≥ 40%)
- Study arm non-equivalence: statistically significant differences between arms in important baseline demographics or risk factors
- Differential retention: (1) significant difference between study arms in characteristics among retained or lost-to-follow up participants; OR (2) more than minimal rate of differential retention (>10%)
- Intervention activities did not match with the intervention concepts or guiding theories intended to produce the desired outcomes
- Did not clearly describe issues related to generalizability
- Too many post hoc analyses (even with Bonferroni corrections)
- Inconsistent findings

All criteria must be satisfied for an intervention to be considered as a best-evidence individual-level, group-level, or couple-level intervention. Source: Lyles et al., (2006) and Lyles at al., (2007).

PRS Efficacy Criteria for Best-Evidence Risk Reduction (RR) Community-level Interventions (CLIs)

Intervention Description

- Clear description of key aspects of the intervention

Quality of Study Design

- Prospective study design
- Appropriate and concurrent control/comparison arm
- ≥ 4 communities per arm or appropriate power analysis indicating that a smaller number of communities was adequate (i.e., 2 or 3 communities per arm)
- Select similar communities (units) for assignment
- To minimize selection bias before assignment regardless of assignment methods (randomization or not); use methods such as systematic, *a priori* approaches to choose intervention and control communities that are similar (e.g., matching or stratification on factors related to important/appropriate community characteristics)

Quality of Study Implementation and Analysis

- Sample individuals from assigned communities in acceptable ways (e.g., random, systematic) and use identical methods and eligibility criteria for selecting participants in each community, study arm, and data collection wave
- If demographic differences are identified *a priori*, differential selection (e.g., over-sampling based on demographics) may be used to achieve equivalence between study arms on those factors
- Follow-up assessment ≥ 3 months post completion of entire time-specific CLI or post full implementation of on-going CLI with recall not referring to pre-intervention period
- "Post full implementation of an on-going CLI" means after all components of the CLI have been started or put in place in communities
- If cohort, at least 70% retention rate at a single follow-up assessment for each study arm
- If cohort chart review, ≥ 70% success rate in matching medical records
- Comparison between intervention arm and an appropriate comparison arm
- Analysis of communities (units) and analysis of individuals within the communities as originally assigned regardless of contamination or logistic/implementation issues
- Analysis of communities (units) regardless of community level of intervention exposure
- Analysis of individuals within the communities (units) regardless of individual level of intervention exposure
- Use of appropriate cluster-level analyses, e.g., adjusting for ICC
- Analysis must be based on post-intervention levels or among pre-post changes in measures
- For pre-post changes used in analysis, measures must be identical, including identical recall period

- Analysis based on an a=.05 (or more stringent) and a 2-sided test; either no statistical differences in baseline levels of the outcome exist or baseline differences are controlled for in the analysis, regardless of allocation method (e.g., randomization, non-randomization)
- No differences on baseline levels of the outcome means reporting no significant difference between groups on BL relevant outcomes or match/stratify/statistically adjust participant data by using propensity scores or relevant outcome covariates (regardless of assignment methods - RCT or non-RCT)

Strength of Evidence

Demonstrated Significant Positive Intervention Effects

- Positive and statistically significant ($p < .05$) intervention effect for ≥ 1 relevant outcome measure
- A positive intervention effect is defined as a greater reduction in HIV/STD incidence or risk behaviors or a greater increase in HIV protective behaviors in the intervention arm relative to the comparison arm.
- A relevant outcome is defined as a behavior (e.g., abstinence, mutual monogamy, number of sex partners, consistent condom use with anal/vaginal sex, unprotected anal/vaginal sex, proportion of anal/vaginal sex acts protected, injection drug use, sharing or borrowing needles/works) that directly impacts HIV risk or a biologic measure indicating HIV or STD infection (i.e., HIV or STD incidence)
- Effect at the follow-up and based on the analyses that meet study implementation and analysis criteria

No Demonstrated Significant Negative Intervention Effects

- No negative and statistically significant ($p < .05$) intervention effect for any relevant outcome
- A negative intervention effect is defined as a greater increase in HIV/STD incidence or risk behaviors or a greater decrease in HIV protective behaviors in the intervention arm relative to the comparison arm.
- No other statistically significant harmful intervention effect
- For an intervention with a replication evaluation, no significant negative intervention effects in the replication study

Additional Limitations to Evaluate

- No evidence that additional limitations resulted in a fatal flaw: o A fatal flaw has occurred when the overall evaluation of limitations resulted in considerable bias, thus substantially reducing the confidence of the findings
- Examples of limitations to check for possible fatal flaw: Group non-equivalence in baseline measures of important demographics or risk factors
- Differential Retention (for cohort studies): (1) association between study arms and characteristics related to retention or attrition; OR (2) more than minimal rate of differential retention (> 10%)
- Differential Refusal: At baseline for cohort studies; by wave for serial cross-sectional studies: (1) association between study arms and characteristics related to refusal; OR (2) more than minimal rate of differential refusal rate (> 100)

- Intervention activities did not match with the intervention concepts or guiding theories intended to produce the desired outcomes
- Did not clearly describe issues related to generalizability
- Effects only found within a potentially biased subset analyses
- Substantial missing data (> 10% or missing data plus loss to attrition does not exceed acceptable limits for retention alone)
- Too many post hoc analyses (even with Bonferroni corrections)
- Pilot study or very small sample size per study arm (< 50)

PRS Efficacy Criteria for Good-Evidence
Risk Reduction (RR) Individual-level, Group-level, and Couple-level Interventions (ILIs/GLIs/CPLs)

Intervention Description

- Clear description of key aspects of the intervention

Quality of Study Design

- Prospective or quasi-prospective study design
- Appropriate and concurrent comparison arm, or historical comparison (provided it is similar to intervention arm with respect to population, setting, and time frame in the epidemic, and identical with respect to follow-up interval, recall period, and outcome measures)
- Random, minimally biased, or moderately biased allocation of participants to study arms, allowing for selection bias unrelated to the intervention or HIV risk. Assignment may be based on pre-established groups or selection into something other than the intervention, provided neither is directly related to HIV risk.

Quality of Study Implementation and Analysis

- Follow-up assessment ≥ 1 month post-completion of intervention for each study arm with recall not referring to pre-intervention period
- At least a 60% retention rate (or medical chart recovery) at a single follow-up for each study arm
- Comparison between intervention arm and an appropriate comparison arm
- Analysis of participants in study arms as originally allocated, or contaminated participants may be excluded if numbers are small, but participants may not be re-assigned for analytic purposes
- Analysis of participants may be based on intervention exposure, where participants exposed to < 50% of the entire intended intervention may be excluded
- If participants excluded due to contamination or low exposure (as described above), retention rate must include these participants at each follow-up they were assessed
- Analysis must be based on post-intervention levels or on pre-post changes in measures
- For pre-post changes used in analysis, measures must be identical, including identical recall period
- Analysis based on an $\alpha = .05$ and either a 2-sided test or 1-sided test if an *a priori* direction is hypothesized
- With nonrandomized assignment, either no statistical differences exist in baseline levels of the outcome measure, or baseline differences must be controlled for in the analysis. If moderately-biased assignment or historical comparison was used, differences in baseline demographics also must be controlled for in the analysis.
- Analytic sample of ≥ 40 participants per study arm

Strength of Evidence
Demonstrated Significant Positive Intervention Effects

- Positive and statistically significant ($p < .05$) intervention effect for ≥ 1 relevant outcome measures

- A positive intervention effect is defined as a greater reduction in HIV/STD incidence or risk behaviors or a greater increase in HIV protective behaviors in the intervention arm relative to the comparison arm.
- A relevant outcome is defined as a behavior (e.g., abstinence, mutual monogamy, number of sex partners, consistent condom use with anal/vaginal sex, unprotected anal/vaginal sex, proportion of anal/vaginal sex acts protected, injection drug use, sharing or borrowing needles/works) that directly impacts HIV risk or a biologic measure indicating HIV or STD infection (i.e., HIV or STD incidence)
- Effect at the follow-up and based on the analyses that meet study implementation and analysis criteria

No Demonstrated Significant Negative Intervention Effects

- No negative and statistically significant ($p < .05$) intervention effect for any relevant outcome
- A negative intervention effect is defined as a greater increase in HIV/STD incidence or risk behaviors or a greater decrease in HIV protective behaviors in the intervention arm relative to the comparison arm.
- No other statistically significant harmful intervention effect
- For an intervention with a replication evaluation, no significant negative intervention effects in the replication study

Additional Limitations to Evaluate

- No evidence that additional limitations resulted in a fatal flaw: o A fatal flaw has occurred when the overall evaluation of limitations indicates they resulted in considerable bias, thus substantially reducing the confidence of the findings.
- Examples of item limitations to check for possible fatal flaw:
    - Effects only found within potentially biased subset analyses
    - Substantial missing data: Missing data plus loss to attrition exceeds acceptable limits for retention alone ($\geq 40\%$)
    - Study arm non-equivalence: statistically significant differences between arms in important baseline demographics or risk factors
    - Differential Retention: (1) association between study arms and characteristics related to retention or attrition; OR (2) more than minimal rate of differential retention ($> 10\%$)
    - Intervention activities did not match with the intervention concepts or guiding theories intended to produce the desired outcomes
    - Did not clearly describe issues related to generalizability
    - Too many post hoc analyses (even with Bonferroni corrections)
    - Inconsistent findings

All criteria must be satisfied for an intervention to be considered as a good-evidence individual-level, group-level, or couple-level intervention

PRS Efficacy Criteria for Good-Evidence Risk Reduction (RR) Community-level Interventions (CLIs)

Intervention Description

- Clear description of key aspects of the intervention

Quality of Study Design

- Prospective or quasi-prospective study design
- Appropriate and concurrent comparison arm, or historical comparison (provided it is similar to intervention arm with respect to population, setting, time frame in the epidemic, and identical with respect to follow-up time, recall period, and outcome measures)
- Post hoc selection of comparison is allowed
- ≥ 1 community per arm
- 1 community per arm is acceptable only if the following conditions are met: (1) there is a significant pre- and post-intervention change in the relevant outcome for the intervention arm, and (2) the significant pre- and post-intervention change is based on appropriate participant-level analysis or repeated-measures analysis.
- Select similar communities (units) for assignment
- To minimize selection bias before assignment regardless of random assignment or other assignment methods, used methods such as systematic, *a priori* approaches to select intervention and comparison communities that are similar (e.g., matching or stratification on factors related to important/appropriate community characteristics)

Quality of Study Implementation and Analysis

- Sample individuals from assigned communities in acceptable ways (e.g., random, systematic) and use identical methods and eligibility criteria for selecting participants in each community, study arm, and data collection wave
- If demographic differences are identified *a priori*, differential selection (e.g., over-sampling based on demographics) may be used to achieve equivalence between study arms on those factors
- Follow-up assessment ≥ 1 month post completion of entire time-specific CLI or post full implementation of on-going CLI with recall not referring to pre-intervention period
- "Post full implementation of on-going CLI" means after all components of the CLI have been started or put in place in communities
- If cohort, at least 60% retention rate (or medical chart recovery) at a single follow-up assessment for each study arm
- Comparison between intervention arm and an appropriate comparison arm
- Analysis of communities (units) as originally assigned, or communities may be excluded due to contamination or logistic/implementation issues only if dropping no more than one community per study arm AND retaining at least two thirds of intended communities
- Analysis of individuals within the communities (units) as originally assigned, or contaminated individuals may be excluded if numbers are small, but individuals may not be reassigned for analytic purposes

- Analysis of communities (units) regardless of community level of intervention exposure
- Analysis of individuals within the communities (units) may be based on intervention exposure, where dropping individuals who were not exposed to any intervention component (e.g., have not heard of or recognized intervention materials) would retain at least 60% of total sample
- Cluster-level analyses may be provided, but is not required
- Analysis must be based on post-intervention levels or among pre-post changes in measures
- For pre-post changes used in analysis, measures must be identical, including identical recall period
- Analysis based on an $\alpha = .05$ and either a 2-sided test or 1-sided test if an a-priori direction is hypothesized
- Either no statistical differences in baseline levels of the outcome exist or baseline differences are controlled for in the analysis, regardless of allocation method (e.g., randomization, non-randomization)
- No differences on baseline levels of the outcome means reporting no significant difference between study arms in baseline relevant outcome measures, or match/stratify/statistically adjust participant data by using propensity scores or relevant outcome covariates (regardless of assignment methods – RCT or non-RCT)

Strength of Evidence

Demonstrated Significant Positive Intervention Effects

- Positive and statistically significant ($p < .05$) intervention effect for ≥ 1 relevant outcome measure
- A positive intervention effect is defined as a greater reduction in HIV/STD incidence or risk behaviors or a greater increase in HIV protective behaviors in the intervention arm relative to the comparison arm
- A relevant outcome is defined as a behavior (e.g., abstinence, mutual monogamy, number of sex partners, consistent condom use with anal/vaginal sex, unprotected anal/vaginal sex, proportion of anal/vaginal sex acts protected, injection drug use, sharing or borrowing needles/works) that directly impacts HIV risk or a biologic measure indicating HIV or STD infection (i.e., HIV or STD incidence)
- Effect at the follow-up and based on the analyses that meet study implementation and analysis criteria

No Demonstrated Significant Negative Intervention Effects

- No negative and statistically significant ($p < .05$) intervention effect for any relevant outcome
- A negative intervention effect is defined as a greater increase in HIV/STD incidence or risk behaviors or a greater decrease in HIV protective behaviors in the intervention arm relative to the comparison arm
- No other statistically significant harmful intervention effect

Additional Limitations to Evaluate

- No evidence that additional limitations resulted in a fatal flaw:

- A fatal flaw has occurred when the overall evaluation of limitations indicate they resulted in considerable bias, thus substantially reducing the confidence of the findings
- Examples of limitations to check for possible fatal flaw:
  - Study arm non-equivalence: statistically significant differences between arms in important baseline demographics or risk factors
  - Differential Retention (for cohort studies): (1) association between study arms and characteristics related to retention or attrition; OR (2) more than minimal rate of differential retention (> 10%)
  - Differential Refusal – at baseline for cohort studies; by wave for serial cross-sectional studies: (1) association between study arms and characteristics related to refusal; OR (2) more than minimal rate of differential refusal rate (> 10%)
  - Intervention activities did not match with the intervention concepts or guiding theories intended to produce the desired outcomes
  - Did not clearly describe issues related to generalizability
  - Effects only found within potentially biased subset analyses
  - Substantial missing data (> 10%, or missing data plus loss to attrition exceeds acceptable limits for retention alone)
  - Too many post hoc analyses (even with Bonferroni corrections)
  - Pilot study or very small sample size per study arm (< 40)
  - Inconsistent findings

Bibliography

Altman, D., Aggleton, P., Williams, M., Kong, T., Reddy, V., Harrad, D.…Parker, R. (2012). Men who have sex with men: Stigma and discrimination. *The Lancet, 380*(9839), 439-445. doi:10.1016/S0140-6736(12)60920-9 [doi]

Baggaley, R. F., White, R. G., & Boily, M. C. (2010). HIV transmission risk through anal intercourse: Systematic review, meta-analysis and implications for HIV prevention. *International Journal of Epidemiology,39*(4), 1048-1063. doi:10.1093/ije/dyq057 [doi]

Baytop, C., Royal, S., Hubbard McCree, D., Simmons, R., Tregerman, R., Robinson, C…Price, C. Comparison of strategies to increase HIV testing among African-American gay, bisexual, and other men who have sex with men in Washington, DC. *AIDS Care, 26*(5), 608-612. doi:10.1080/09540121.2013.845280

Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101.

Beyrer, C., Baral, S. D., van Griensven, F., Goodreau, S. M., Chariyalertsak, S., Wirtz, A. L., Brookmeyer, R. (2012). Global epidemiology of HIV infection in men who have sex with men. *The Lancet, 380*(9839), 367-377. doi:10.1016/S0140-6736(12)60821-6

Boily, M. C., Baggaley, R. F., Wang, L., Masse, B., White, R. G., Hayes, R. J., Alary, M. (2009). Heterosexual risk of HIV-1 infection per sexual act: Systematic

review and meta-analysis of observational studies. *The Lancet, 9*(2), 118-129. doi:10.1016/S1473-3099(09)70021-0

Bollen, C. W., Uiterwaal, C. S., & van Vught, A. J. (2003). Cumulative metaanalysis of high-frequency versus conventional ventilation in premature neonates. *Am J Respir Crit Care Med, 168*(10), 1150-1155. doi: 10.1164/rccm.200306-721CP

Bornstein, M., Hedges, L., Higgins, J.P.T., & Rothstein, H. (2009). *Introduction to meta-analysis.* West Sussex, United Kingdom: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J., Rothstein, H. R. (2015). Biostat: Comprehensive Meta-analysis (Version 3) [Software]. Available from http://www.meta-analysis.com

Borenstein, M., Hedges, L.V., Higgins, J., Rothstein, H.R.(2015). Comprehensive Meta-analysis: Version 3 (User Manual). Retrieved from: https://www.meta-analysis.com/downloads/Meta-Analysis%20Manual%20V3.pdf. Last accessed March 2, 2016.

Centers for Disease Control and Prevention (CDC). (2011). HIV testing among men who have sex with men--21 cities, United States, 2008. *MMWR Morbidity and Mortality Weekly Report, 60*(21), 694-699. doi:mm6021a3

Centers for Disease Control. (2012). Estimated HIV incidence in the united states, 2007–2010. *HIV Surveillance Supplemental Report, 17*(4). Retrieved from

http://www.cdc.gov/hiv/pdf/statistics_hssr_vol_17_no_4.pdf. Last accessed November 9, 2014.

Centers for Disease Control. (2013a). Diagnoses of HIV infection in the United States and dependent areas, 2011. *HIV Surveillance Report, 23.* Retrieved from *http://www.cdc.gov/hiv/pdf/statistics_2011_HIV_Surveillance_Report_vol_23.pdf* . Last accessed November 9, 2014.

Centers for Disease Control and Prevention (CDC). (2013b). HIV testing and risk behaviors among gay, bisexual, and other men who have sex with men - United States. *MMWR Morbidity and Mortality Weekly Report, 62*(47), 958-962. doi:mm6247a4

Centers for Disease Control (2014a). CDC fact sheet. HIV among gay and bisexual men. Retrieved from http://www.cdc.gov/hiv/risk/gender/msm/facts/. Last accessed November 9, 2014.

Centers for Disease Control. (2014b). Compendium of evidence-based interventions and best practices for HIV prevention. Retrieved from http://www.cdc.gov/hiv/prevention/research/compendium/rr/complete.html. Last accessed November 9, 2014.

Centers for Disease Control (2014c). HIV Risk Reduction Evidence Review (Efficacy Criteria). Retrieved from http://www.cdc.gov/hiv/dhap/prb/prs/efficacy/rr/criteria/. Last accessed March 1, 2016.

Centers for Disease Control (2015a). High-Impact Agenda. Retrieved from: http://www.cdc.gov/hiv/policies/hip/hip.html Last accessed March 3, 2016.

Centers for Disease Control (2015b). PRS RR Priortization Plan. Retrieved from: http://www.cdc.gov/hiv/pdf/prs/PRS%20RR%20Prioritization%20Plan.pdf Last accessed March 2, 2016.

Center for Evidence-Based Research. OCENM Levels of Evidence. Retrieved from http://www.cebm.net/ocebm-levels-of-evidence/ Last accessed February 2, 2015.

Chesney, M. A., Koblin, B. A., Barresi, P. J., Husnik, M. J., Celum, C. L., Colfax, G., … the EXPLORE Study Team. An Individually Tailored Intervention for HIV Prevention: Baseline Data From the EXPLORE Study. *American Journal of Public Health, 93*(6), 933–938.

Clarke, M., Brice, A., & Chalmers, I. (2014). Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS One, 9*(7), e102670. doi: 10.1371/journal.pone.0102670

Coates, T. J., Richter, L., & Caceres, C. (2008). Behavioral strategies to reduce HIV transmission: How to make them work better. *Lancet, 372*(9639), 669-684. doi:10.1016/S0140-6736(08)60886-7

Coffin, P. O., Santos, G. M., Colfax, G., Das, M., Matheson, T., DeMicco, E…Herbst, J.H. (2014). Adapted personalized cognitive counseling for episodic substance-using men who have sex with men: A randomized controlled trial. *AIDS and Behavior, 18*(7), 1390-1400. doi:10.1007/s10461-014-0712-4

Cohen, M. S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N…Fleming, T.R. (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *The New England Journal of Medicine,365*(6), 493-505. doi:10.1056/NEJMoa1105243

Cohen, M.S., Smith, M.K., Muessig, K.E., Hallett, T.B., Powers, K.A., & Kashuba, A.D. (2013). Antiretroviral treatment of HIV-1 prevents transmission of HIV-1: Where do we go from here? *Lancet,382*(9903), 1515-1524. doi:10.1016/S0140-6736(13)61998-4

Collins, C.B. (2015). Prioritization, policy, and integration – The course of dissemination of evidence-based HIV behavioral interventions in the United States *National HIV Prevention Conference*. Atlanta: GA.

Cook, T.D., Cooper, H., Cordray, D., Hartmann, H., Hedges, L.V., Light, R., Louis, T., Cordray, D.S. (1992). *Meta-analysis for Explanation.* New York: Russell Sage Foundation.

Crepaz, N., Lyles, C. M., Wolitski, R. J., Passin, W. F., Rama, S. M., Herbst, J. H., . . . Stall, R. (2006). Do prevention interventions reduce HIV risk behaviours among people living with HIV? A meta-analytic review of controlled trials. *AIDS, 20*(2), 143-157. doi: 10.1097/01.aids.0000196166.48518.a0

Das, M., Chu, P.L., Santos, G.M., Scheer, S., Vittinghoff, E., McFarland, W., Colfax, G.N. (2010). Decreases in community viral load are accompanied by reductions in new HIV infections in san Francisco. *PloS One,5*(6), e11068. doi:10.1371/journal.pone.0011068 [doi]

Dickersin, K. & Min, Y.I. (1993). NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials, Doc No. 50.*

Dieffenbach, C.W., & Fauci, A.S. (2011). Thirty years of HIV and AIDS: Future challenges and opportunities. *Annals of Internal Medicine, 154*(11), 766-771. doi:10.7326/0003-4819-154-11-201106070-00345 [doi]

Dilley, J. W., Woods, W.J., Loch, L., Nelson, K., Sheon, N., Mullan, J…McFarland, W. (2007) Brief cognitive counseling with HIV testing to reduce sexual risk among men who have sex with men: results from a randomized controlled trial using paraprofessional counselors. *Journal of Acquired Immune Deficiency Syndromes* 44, 569-577 DOI: 10.1097/QAI.0b013e318033ffbd

Donner, A., Koval, J. J (1980). The Estimation of Intraclass Correlation in the Analysis of Family Data. *Biometrics, 36*(1),19–25.doi:10.2307/2530491

Du Bois, S. N., & McKirnan, D. J. (2012). A longitudinal analysis of HIV treatment adherence among men who have sex with men: A cognitive escape perspective. *AIDS Care, 24*(11), 1425-1431. doi:10.1080/09540121.2011.650676

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Ekkekakis, P. (2015). Honey, I shrunk the pooled SMD! Guide to critical appraisal of systematic reviews and meta-analyses using the Cochrane review on exercise for depression as example. *Mental Health & Physical Activity, 8,*21-36.

Ellen, J.M., McCree, D.H., Muvva, R., Chung, S.E., Miazad, R.M., Arrington-Sanders, R….Fitchtenberg, C. (2013). Recruitment approaches to identifying newly diagnosed HIV infection among African-American men who have sex with men. *International Journal of STD & AIDS, 24*(5), 335-339. doi:10.1177/0956462412472459 [doi]

Flores, S.A. & Crepaz, N. (2004). Quality of Study methods in individual- and group-level HIV intervention research: critical reporting elements. *AIDS Education and Prevention, 16*(4), 341-352.

Frieden TR, Foti KE, Mermin J. (2015). Applying public health principles to the HIV epidemic -- How are we doing? *New England Journal of Medicine, 373*(23), 2281-2287.

Fuqua, V., Chen, Y.H., Packer, T., Dowling, T., Ick, T.O., Nguyen, B….Raymond, H.F. (2012). Using social networks to reach black MSM for HIV testing and linkage to care. *AIDS and Behavior, 16*(2), 256-265. doi:10.1007/s10461-011-9918-x

Gleser, L. J., & Olkin, I. (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine, 15*(23), 2493-2507. doi: 10.1002/(sici)1097-0258(19961215)15:23<2493::aid-sim381>3.0.co;2-c

Grant, R.M., Lama, J.R., Anderson, P.L., McMahan, V., Liu, A.Y., Vargas, L.,…Glidden, D.F. (2010). Pre-exposure chemoprophylaxis for HIV prevention in men who have sex with men. *The New England Journal of Medicine, 363*(27), 2587-2599. doi:10.1056/NEJMoa1011205

Grossman, C. I., Purcell, D. W., Rotheram-Borus, M. J., & Veniegas, R. (2013). Opportunities for HIV Combination Prevention to Reduce Racial and Ethnic Health Disparities. *The American Psychologist*, *68*(4), 237–246. http://doi.org/10.1037/a0032711

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA, 274*(22), 1800-1804.

Haddock, C., Rindskopf, D., & Shadish, W. (1988). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339-353.

Halkitis, P.N., Kupprat, S.A., McCree, D.H., Simons, S.M., Jabouin, R., Hampton, M.C., Gillen, S. (2011). Evaluation of the relative effectiveness of three HIV testing strategies targeting African-American men who have sex with men (MSM) in New York City. *Annals of Behavioral Medicine, 42*(3), 361-369. doi:10.1007/s12160-011-9299-4

Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: application to programs for reducing sexual violence. *Sex Abuse, 17*(4), 357-373. doi: 10.1007/s11194-005-8049-1

Harawa, N. T., Williams, J. K., McCuller, W. J., Ramamurthi, H. C., Lee, M., Shapiro, M. F., . . . Cunningham, W. E. (2013). Efficacy of a culturally congruent HIV risk-reduction intervention for behaviorally bisexual black men: results of a randomized trial. *AIDS, 27*(12), 1979-1988.

Hatfield, L. A., Ghiselli, M. E., Jacoby, S. M., Cain-Nielsen, A., Kilian, G., McKay, T., & Rosser, B. R. (2010). Methods for recruiting men of color who have sex with men in prevention-for-positives interventions. *Prev Sci, 11*(1), 56-66. doi: 10.1007/s11121-009-0149-6

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-108.

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, *96*, 573–580. doi:10.1037/0033-2909.96.3.573

Herbst, J.H., Beeker, C., Mathew, A., McNally, T., Passin, W.F., Kay, L.S…Johnson, R.L. (2007). The effectiveness of individual-, group-, and community-level HIV behavioral risk-reduction interventions for adult men who have sex with men: A systematic review. *American Journal of Preventive Medicine, 32*(4 Suppl), S38-67. doi:S0749-3797(06)00555-1

Herbst, J.H., Sherba, R.T., Crepaz, N., Deluca, J.B., Zohrabyan, L., Stall, R.D., Lyles, C.M. (2005). A meta-analytic review of HIV behavioral interventions for reducing sexual risk behavior of men who have sex with men. *Journal of Acquired Immune Deficiency Syndromes*, *39*(2), 228-241. doi:00126334-200506010-00016

Higa, D.H., Crepaz, N., Marshall K.J., Kay, L., Vosburgh, W., Spikes, P….Purcell, D.W. (2013). A systematic review to identify challenges of demonstrating efficacy of HIV behavioral interventions for gay, bisexual, and other men who have sex with men. *AIDS & Behavior, 17*(4), 12-31-1244. doi: 10.1007/s10461-013-0418-z.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*,*327*(7414), 557–560.

Higgins, J.P.T. & Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions. West Sussex, United Kingdom: John Wiley & Sons Ltd.

Hirshfield, S., Chiasson, M. A., Joseph, H., Scheinmann, R., Johnson, W. D., Remien, R. H., . . . Margolis, A. D. (2012). An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. *PLoS One, 7*(10), e46252. doi:10.1371/journal.pone.0046252

Ioannidis, J. P., Contopoulos-Ioannidis, D. G., & Lau, J. (1999). Recursive cumulative meta-analysis: a diagnostic for the evolution of total randomized evidence from group and individual patient data. *J Clin Epidemiol, 52*(4), 281-291.

Ioannidis, J. P., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *BMJ, 335*(7626), 914-916. doi: 10.1136/bmj.39343.408449.80

Institute of Medicine (2001). *No time to lose: Getting more from HIV prevention.* Washington, DC: National Academies Press.

Johnson BT, Redding CA, DiClemente RJ, Dodge BM, Mustanski BS, Sheeran P, Warren MR, Zimmerman RS, Fisher WA, Conner MT, Carey MP, Fisher, JD, Stall RD, Fishbein M. (2010). A Network-Individual-Resource model for HIV prevention. *AIDS & Behavior*, 14 (Supp. 2)**,** S204-S221.

Johnson, W. D., Diaz, R. M., Flanders, W. D., Goodman, M., Hill, A. N., Holtgrave, D...McClellan, W.M. (2008). Behavioral interventions to reduce risk for sexual transmission of HIV among men who have sex with men. *The Cochrane Database of Systematic Reviews, (3):CD001230. doi*(3), CD001230. doi:10.1002/14651858.CD001230.pub2 [doi]

Johnson, W.D., Hedges, L.V., Ramirez, G., Semaan, S., Norman, L.R., Sogolow, E., Diaz, R.M (2002a). HIV prevention research for men who have sex with men: A systematic review and meta-analysis. *Journal of Acquired Immune Deficiency Syndromes (1999), 30 Suppl 1*, S118-29.

Johnson, W.D., Semaan, S., Hedges, L.V., Ramirez, G., Mullen, P.D., & Sogolow, E. (2002b). A protocol for the analytical aspects of a systematic review of HIV prevention research. *Journal of Acquired Immune Deficiency Syndromes (1999), 30 Suppl 1*, S62-72.

Johnson, W.D., Holtgrave, D.R., McClellan, W.M., Flanders, W.D., Hill, A.N., & Goodman, M. (2005). HIV intervention research for men who have sex with men: A 7-year update. *AIDS Education and Prevention, 17*(6), 568-589. doi:10.1521/aeap.2005.17.6.568

Joint United Nations Program on HIV/AIDS. (2010). Combination HIV prevention: tailoring and coordinating biomedical, behavioral and structural strategies 10 to reduce new HIV infections A UNAIDS discussion paper. Retrieved from http://www.unaids.org/sites/default/files/media_asset/JC2007_Combination_Prevention_paper_en_0.pdf. Last accessed November 16, 2014.

Kaufman, M.R., Cornish, F., Zimmerman, R.S., & Johnson, B.T. (2014). Health behavior change models for HIV prevention and AIDS care: Practical recommendations for a multi-level approach. *Journal of Acquired Immune Deficiency Syndromes (1999), 66 Suppl 3*, S250-8. doi:10.1097/QAI.0000000000000236

Kegeles, S. M., Hays, R. B., Pollack, L. M., & Coates, T. J. (1999). Mobilizing young gay and bisexual men for HIV prevention: a two-community study. *AIDS, 13*(13), 1753-1762.

Kelly, J. A., St Lawrence, J. S., Diaz, Y. E., Stevenson, L. Y., Hauth, A. C., Brasfield, T. L., . . . Andrew, M. E. (1991). HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. *American Journal of Public Health*, *81*(2), 168-171.

Kelly, J. A., Murphy, D. A., Sikkema, K. J., McAuliffe, T. L., Roffman, R. A., Solomon, L. J., . . . Kalichman, S. C. (1997). Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. Community HIV Prevention Research Collaborative. *Lancet, 350*(9090), 1500-1505

Klein, J. B., Jacobs, R. H., & Reinecke, M. A. (2007). Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. *J Am Acad Child Adolesc Psychiatry*, *46*(11), 1403-1413. doi: 10.1097/chi.0b013e3180592aaa

Kippax S. (2003). Sexual health interventions are unsuitable for experimental evaluation. In Stephenson, J., Imrie, J, & Bonnell, C. (Ed.), *Effective Sexual Health Interventions: Issues in Experimental Evaluation* (pp. 17-34). New York: Oxford University Press.

Koblin, B. A., Chesney, M., & Coates, T. (2004). Effects of a behavioural intervention to reduce acquisition of HIV infection among men who have sex with men: The EXPLORE randomised controlled study. *The Lancet*, *364*(9428), 41-50. doi: 10.1016/S0140-6736(04)16588-4

Koblin, B.A., Husnik, M.J., Colfax, G., Huang, Y., Madison, M., Mayer, K…Buchbinder, S. (2006). Risk factors for HIV infection among men who have sex with men. *AIDS, 20*(5), 731-739. doi:10.1097/01.aids.0000216374.61442.55

Kurth, A.E., Celum, C., Baeten, J.M., Vermund, S.H., & Wasserheit, J.N. (2011). Combination HIV prevention: Significance, challenges, and opportunities. *Current HIV/AIDS Reports, 8*(1), 62-72. doi:10.1007/s11904-010-0063-3

Kurtz, S.P., Stall, R.D., Buttram, M.E., Surratt, H.L., & Chen, M. (2013). A randomized trial of a behavioral intervention for high risk substance-using MSM. *AIDS and Behavior, 17*(9), 2914-2926. doi:10.1007/s10461-013-0531-z

Lyles C.M., Crepaz, N., Herbst, J.H., Kay, L.S. (2006). Evidence-based HIV behavioral prevention from the perspective of the CDC's HIV/AIDS Prevention Research Synthesis Team. *AIDS Education and Prevention, 18*(4 Suppl A): 21-31. doi: 10.1521/aeap.2006.18.supp.21

Maddock, J. E., & Rossi, J. S. (2001). Statistical power of articles published in three health psychology-related journals. *Health Psychol, 20*(1), 76-78.

Malta, M., Magnanini, M.M., Strathdee, S.A., & Bastos, F. I. (2010). Adherence to antiretroviral therapy among HIV-infected drug users: A meta-analysis. *AIDS and Behavior, 14*(4), 731-747. doi:10.1007/s10461-008-9489-7

Mansergh, G., Koblin, B.A., McKirnan, D.J., Hudson, S.M., Flores, S.A., Wiegand, R.E…Colfax, G.N. (2010). An intervention to reduce HIV risk behavior of substance-using men who have sex with men: A two-group randomized trial with

a nonrandomized third group. *PLoS Medicine, 7*(8), e1000329. doi:10.1371/journal.pmed.1000329 [doi]

Marks, G., Crepaz, N., & Janssen, R.S. (2006). Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *AIDS, 20*(10), 1447-1450. doi:10.1097/01.aids.0000233579.79714.8d [doi]

Marsh, K.L., Johnson, B.T., & Carey, M.P. (2001). Conducting meta-analyses of HIV prevention literatures from a theory-testing perspective. *Evaluation & the Health Professions, 24*(3), 255-276.

Maulsby, C., Millett, G., Lindsey, K., Kelley, R., Johnson, K., Montoya, D., Holtgrave, D. (2013). A systematic review of HIV interventions for black men who have sex with men (MSM). *BMC Public Health, 13*, 625-2458-13-625. doi:10.1186/1471-2458-13-625 [doi]

Mayer, K.H., Bekker, L.G., Stall, R., Grulich, A.E., Colfax, G., & Lama, J.R. (2012). Comprehensive clinical care for men who have sex with men: An integrated approach. *Lancet, 380*(9839), 378-387. doi:10.1016/S0140-6736(12)60835-6

Mayer, K.H., Wang, L., Koblin, B., Mannheimer, S., Magnus, M., del Rio, C.…Eshelmen, S.H. (2014). Concomitant socioeconomic, behavioral, and biological factors associated with the disproportionate HIV infection burden among black men who have sex with men in 6 U.S. cities. *PloS One, 9*(1), e87298. doi:10.1371/journal.pone.0087298 [doi]

McCree, D.H., Millett, G., Baytop, C., Royal, S., Ellen, J., Halkitis, P…Gillen, S. (2013). Lessons learned from use of social network strategy in HIV testing

programs targeting African American men who have sex with men. *American Journal of Public Health, 103*(10), 1851-1856. doi:10.2105/AJPH.2013.301260 [doi]

Melendez-Torres, G., & Bonell, C. (2013). Systematic review of cognitive behavioral interventions for HIV risk reduction in substance-using men who have sex with men. *International Journal of STD & AIDS,25*(9), 627-635. doi:0956462413515638

Menza, T. W., Jameson, D. R., Hughes, J. P., Colfax, G. N., Shoptaw, S., & Golden, M. R. (2010). Contingency management to reduce methamphetamine use and sexual risk among men who have sex with men: a randomized controlled trial. *BMC Public Health, 10*, 774. doi:10.1186/1471-2458-10-774

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed.1000097

Morgenstern, J., Kuerbis, A.N., Chen, A.C., Kahler, C.W., Bux, D. A., & Kranzler, H.R. (2012). A randomized clinical trial of naltrexone and behavioral therapy for problem drinking men who have sex with men. *Journal of Consulting and Clinical Psychology, 80*(5), 863-875. doi:2012-12953-001 [pii]

Murad, M.H., Montori, V., Ioannidis, J.A., Jaeschke, R., Devereaux, P.R., Kameshwar, P…Guyatt, G. (2014). How to read a systematic review and meta-analysis and apply the results to patient care users' guides to the medical literature. *Journal of the American Medical Association, 312*(2), 171-179.

Noar, S. M. (2008). Behavioral interventions to reduce HIV-related sexual risk behavior: Review and synthesis of meta-analytic evidence. *AIDS and Behavior, 12*(3), 335-353. doi:10.1007/s10461-007-9313-9 [doi]

O'Donnell, L., Stueve, A., Joseph, H. A., & Flores, S. (2014). Adapting the VOICES HIV behavioral intervention for Latino men who have sex with men. *AIDS Behav, 18*(4), 767-775. doi: 10.1007/s10461-013-0653-3

Orwin, R.G. & Boruch, R.F. (1983). RRT meets RDD: statistical strategies for assuring response privacy in telephone surveys. *Public Opinion Quarterly, 46,* 56-571.

Ostrow, D. E., Fox, K. J., Chmiel, J. S., Silvestre, A., Visscher, B. R., Vanable, P. A., . . . Strathdee, S. A. (2002). Attitudes towards highly active antiretroviral therapy are associated with sexual risk taking among HIV-infected and uninfected homosexual men. *AIDS, 16*(5), 775-780.

Ostrow, D.G. Stall, R. (2008) Alcohol, tobacco, and drug use among gay and bisexual men. In Wolitski, R.J., Stall, R., Valdiserri, R.O., Unequal Opportunity: Health Disparities Affecting Gay and Bisexual Men in the United States. New York: Oxford University Press.

Parsons, J. T., Grov, C., & Golub, S. A. (2012). Sexual compulsivity, co-occurring psychosocial health problems, and HIV risk among gay and bisexual men: further evidence of a syndemic. *American Journal of Public Health, 102*(1), 156-162. doi: 10.2105/ajph.2011.300284

Parsons, J. T., Lelutiu-Weinberger, C., Botsko, M., & Golub, S. A. (2014). A randomized controlled trial utilizing motivational interviewing to reduce HIV risk and drug use in young gay and bisexual men. *Journal of Consulting and Clinical Psycholology, 82*(1), 9-18. doi: 10.1037/a0035311

Patel, P., Borkowf, C. B., Brooks, J. T., Lasry, A., Lansky, A., & Mermin, J. (2014). Estimating per-act HIV transmission risk: a systematic review. *AIDS, 28*(10), 1509-1519. doi: 10.1097/qad.0000000000000298

Paz-Bailey, G., Hall, H.I., Wolitski, R.J., Prejean, J., Van Handel, M.M. Le, B…. Valleroy, L.A. HIV Testing and Risk Behaviors Among Gay, Bisexual, and Other Men Who Have Sex with Men — United States. *Morbidity and Mortality Weekly Report, 62*(47), 958-962.

Pham, B., Platt, R., McAuley, L., Sampson, M., Klassen, T., & Moher, D. (2001). Detecting and minimizing the impact of for publication bias: An empirical study of methods. *Evaluation and the Health Professions, 24*(2), 109-125.

Picciano, J. F., Roffman, R. A., Kalichman, S. C., & Walker, D. D. (2007). Lowering obstacles to HIV prevention services: effects of a brief, telephone-based intervention using motivational enhancement therapy. *Annals of Behavioral Medicine, 34*(2), 177-187. doi:10.1080/08836610701566894

Purcell, D.W., Johnson, C.H., Lansky, A., Prejean, J., Stein, R., Denning, P…Crepaz, N. (2012). Estimating the population size of men who have sex with men in the

United States to obtain HIV and syphilis rates. *The Open AIDS Journal, 6,* 98-107. doi:10.2174/1874613601206010098

Rhodes, S. D., Vissman, A.T., Stowers, J., Miller, C., McCoy, T.P., Hergenrather, K.C…Eng, E. (2011). A CBPR partnership increases HIV testing among men who have sex with men (MSM): Outcome findings from a pilot test of the CyBER/testing internet intervention. *Health Education & Behavior, 38*(3), 311-320. doi:10.1177/1090198110379572

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3),638-641.http://dx.doi.org/10.1037/0033-2909.86.3.638

Ross, D.A. & Wight, D. (2003). In Stephenson, J., Imrie, J, & Bonnell, C. (Ed.), The role of randomized controlled trials in assessing sexual health interventions. *Effective Sexual Health Interventions: Issues in Experimental Evaluation* (pp. 35-50). New York: Oxford University Press.

Rossi, J. S. (1990). Statistical power of psychological research: what have we gained in 20 years? *J Consult Clin Psychol, 58*(5), 646-656.

Santos, G. M., Coffin, P. O., Vittinghoff, E., DeMicco, E., Das, M., Matheson, T…Dilley, J.W. (2014). Substance use and drinking outcomes in personalized cognitive counseling randomized trial for episodic substance-using men who have sex with men. *Drug and Alcohol Dependence, 138*, 234-239. doi:10.1016/j.drugalcdep.2014.02.015 [doi]

Schnall, R., Travers, J., Rojas, M., & Carballo-Dieguez, A. (2014). eHealth interventions for HIV prevention in high-risk men who have sex with men: A systematic review. *Journal of Medical Internet Research,16*(5), e134. doi:10.2196/jmir.3393

Schulz, K. F., Altman, D. G., & Moher, D. (2011). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Int J Surg*, *9*(8), 672-677. doi: 10.1016/j.ijsu.2011.09.004

Schwarcz, S.K., Chen, Y.H., Murphy, J. L., Paul, J. P., Skinta, M.D., Scheer, S….Dilley, J.W. (2013). A randomized control trial of personalized cognitive counseling to reduce sexual risk among HIV-infected men who have sex with men. *AIDS Care*, *25*(1), 1-10. doi:10.1080/09540121.2012.674095

Singh, S., Bradley, H., Hu, X., Skarbinski, J., Hall, H.I., & Lansky, A. (2014). Men living with diagnosed HIV who have sex with men: Progress along the continuum of HIV care - United States, 2010.*MMWR.Morbidity and Mortality Weekly Report*, *63*(38), 829-833. doi:mm6338a2

Singer, M. (2010). Pathogen-pathogen interaction: a syndemic model of complex biosocial processes in disease. *Virulence*, *1*(1), 10-18. doi: 10.4161/viru.1.1.9933

Shoptaw, S., Reback, C. J., Peck, J. A., Yang, X., Rotheram-Fuller, E., Larkins, S., . . . Hucks-Ortiz, C. (2005). Behavioral treatment approaches for methamphetamine dependence and HIV-related sexual risk behaviors among urban gay and bisexual men. *Drug and Alcohol Dependence*, *78*(2), 125-134.

Stall, R., Duran, L., Wisniewski, S. R., Friedman, M. S., Marshal, M. P., McFarland, W., … Mills, T. C. (2009). Running in Place: Implications of HIV Incidence Estimates among Urban Men Who Have Sex with Men in the United States and Other Industrialized Countries. *AIDS and Behavior*, *13*(4), 615–629. http://doi.org/10.1007/s10461-008-9509-7

Starks, T. J., Millar, B. M., Eggleston, J. J., & Parsons, J. T. (2014). Syndemic factors associated with HIV risk for gay and bisexual men: comparing latent class and latent factor modeling. *AIDS Behav, 18*(11), 2075-2079. doi: 10.1007/s10461-014-0841-9

Stephenson, J. & Irmie, J. (1998). Why do we need randomized controlled trials to assess behavioral interventions? *BMJ*, 316:611-613.

Stockman, J. K., Schwarcz, S. K., Butler, L. M., de Jong, B., Chen, S. Y., Delgado, V., & McFarland, W. (2004). HIV prevention fatigue among high-risk populations in San Francisco. *J Acquir Immune Defic Syndr*, *35*(4), 432-434.

Sullivan, P. S., Carballo-Diéguez, A., Coates, T., Goodreau, S. M., McGowan, I., Sanders, E. J., … Sanchez, J. (2012). Successes and challenges of HIV prevention in men who have sex with men. *Lancet*, *380*(9839), 388–399. http://doi.org/10.1016/S0140-6736(12)60955-6

The Guide to Community Preventive Services (2014). Data Abstraction Form. Retrieved from http://www.thecommunityguide.org/methods/abstractionform.pdf Last accessed August 7, 2015.

The White House. (July 2010). National HIV/AIDS strategy for the United States. Retrieved from http://www.whitehouse.gov/sites/default/files/uploads/NHAS.pdf. Last accessed November 16, 2014.

Trikalinos, T. A., Churchill, R., Ferri, M., Leucht, S., Tuunainen, A., Wahlbeck, K., & Ioannidis, J. P. (2004). Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol, 57*(11), 1124-1130.

Turner, R. M., Bird, S. M., & Higgins, J. P. (2013). The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One, 8*(3), e59202. doi: 10.1371/journal.pone.0059202

U.S. Department of State. (2013). PEPFAR blueprint: Creating an AIDS-free generation. Retrieved from http://www.pepfar.gov/documents/organization/201386.pdf. Last accessed November 16, 2014.

Wilton, L., Herbst, J. H., Coury-Doniger, P., Painter, T. M., English, G., Alvarez, M. E., . . . Carey, J. W. (2009). Efficacy of an HIV/STI prevention intervention for black men who have sex with men: findings from the Many Men, Many Voices (3MV) project. *AIDS and behavior*, *13*(3), 532-544. doi:10.1007/s10461-009-9529-y

van den Berg, J., Larson, H., Zimet, G., & Lally, M. (2014). Bundling human papillomavirus vaccination and rapid human immunodeficiency virus testing for young gay and bisexual men. *LGBT Health, 1*(3), 233-237.

van den Berg JJ, Dong K, Tang A, Wanke CA, Larson E, Robinette AE, Nwaoha F, Lally MA. Examining the relationships between food insecurity, depression, and HIV risk behaviors among substance users who are homeless in the northeastern United States. Poster presentation at: 2015 National HIV Prevention Conference, December 6-9, 2015; Atlanta, GA. Abstract #2170.

Velasquez, M. M., von Sternberg, K., Johnson, D.H., Green, C., Carbonari, J.P., & Parsons, J.T. (2009). Reducing sexual risk behaviors and alcohol use among HIV-positive men who have sex with men: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *77*(4), 657-667. doi:10.1037/a0015519 [doi]

Vittinghoff, E., Douglas, J., Judson, F., McKirnan, D., MacQueen, K., & Buchbinder, S.P. (1999). Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *American Journal of Epidemiology*, *150*(3), 306-311.

Wejnert, C., Le, B., Rose, C.E., Oster, A.M., Smith, A.J., & Zhu, J. (2013). HIV infection and awareness among men who have sex with men-20 cities, United States, 2008 and 2011. *PloS One, 8*(10), e76878. doi:10.1371/journal.pone.0076878 [doi]

White, J.M., Gordon, J.R., & Mimiaga, M.J. (2014). The role of substance use and mental health problems in medication adherence among HIV-infected MSM. *LGBT Health,* epub ahead of print, retrieved from http://online.liebertpub.com/doi/abs/10.1089/lgbt.2014.0020?journalCode=lgbt

Ye, S., Yin, L., Amico, R., Simoni, J., Vermund, S., Ruan, Y., … Qian, H.-Z. (2014).

Efficacy of Peer-Led Interventions to Reduce Unprotected Anal Intercourse among Men Who Have Sex with Men: A Meta-Analysis. *PLoS ONE*, *9*(3), e90788. http://doi.org/10.1371/journal.pone.0090788

Zhang, W., Nuki, G., Moskowitz, R. W., Abramson, S., Altman, R. D., Arden, N. K., . . . Tugwell, P. (2010). OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage*, *18*(4), 476-499. doi: 10.1016/j.joca.2010.01.013